1991

Probabilistic Models Of Search State And Path Patterns In Hypertext Information Retrieval Systems

Liwen Qiu

Probabilistic Models of Search State and Path Patterns in Hypertext Information Retrieval Systems

By

Liwen Qiu

School of Library and Information Science

Submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario

June 1991

ISBN 0-315-66285-9

ABSTRACT

The objective of this research is to discover the search state and path patterns through which users retrieve information in hypertext systems. The Markov model is used to describe users' search behaviour. As determined by the log-linear model test, the second-order Markov model is the best model. Search patterns of different user groups were studied by comparing the corresponding transition probability matrices. The comparisons were made based on the following factors: gender, search experience, search task, and the user's academic background. The statistical tests revealed that there were significant differences among all the groups being compared.

A three-way analysis of variance test was conducted to study the effects of gender, search task, and search experience on search option (analytical vs. browsing), as measured by the proportion of nodes reached through analytical searching. The search task factor influenced search option in that a general task caused more browsing and a specific task more analytical searching. Search experience alone did not affect the search option. There was a possible gender difference among high experience users, with males favouring analytical searching more than females.

Two frequency distribution models were developed and tested to describe path patterns. Path length followed a shifted negative binomial distribution. The frequency of node visiting followed a Zipf distribution.

The resulting probabilistic models can help us better understand users' search behaviour and the search process involved. They provide valuable information in evaluating a system's existing operation and in refining future design. They also provide a background for examination of systems via simulation studies.

ACKNOWLEDGEMENT

I am deeply grateful to the members of my advisory committee, Dr. P. T. Nicholls, Dr. M. J. Nelson, and especially my chief supervisor Dr. J. M. Tague-Sutcliffe. The dissertation certainly owes a great deal to their advice, guidance, comments and encouragement. A special thanks to Dr. M. T. Kinnucan who willingly gave his time and energy to discuss the research and provide valuable suggestions. I am also grateful to Mr. D. S. Vaughan for all his help and support during the research and the writing. Last but not least, I would like to express my sincere appreciation to everybody who kindly participated in the experiments, and Dr. F. T. Dolan who supported the experiment. Without their cooperation, this dissertation would have been impossible.

TABLE OF CONTENTS

Page

CERTIFICATE OF EXAMINATION ii

ABSTRACT iii

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vi

LIST OF TABLES ix

LIST OF FIGURES xi

LIST OF APPENDICES xii

Chapter 1 -- Introduction 1

Chapter 2 -- Basic Knowledge of Markov Models 8

Chapter 3 -- Literature Review 12

3.1 Application of Markov Models in Library and Information Science 12

3.1.1 Markov Model in the Information Retrieval Systems 13

3.1.2 Markov Model in Library Circulation Process 21

3.1.3 Bibliometric Application of Markov Model ... 23

3.2 Research in Hypertext Information Retrieval Systems 25

3.3 User Studies in Information Retrieval Systems .. 32

3.4 Probabilistic Models for Traditional Information Retrieval Systems 34

Chapter 4 -- Methodology 38

4.1 The Experimental System 39

4.2 Subjects 49

4.3 Types of Search Tasks 52

4.4 Data Collection 53

4.5 Data Processing 55

4.5.1 Distribution Models 56

4.5.2 The Markov Model 56

4.5.3 Analysis of Variance 64

Chapter 5 -- Results 68

5.1 The Markov Model 68

5.2 Comparison of Different User Groups Using Markov Models 78

5.2.1 Gender 79

5.2.2 Task 83

5.2.3 Search Experience 89

5.2.4 Subject's Academic Background 93

5.3 Comparison of Search Options (analytical searching vs. browsing) 98

5.4 Path Length Distribution 105

5.4.1 Shifted Binomial Distribution 107

5.4.2 Shifted Poisson Distribution 107

5.4.3 Shifted Negative Binomial Distribution 108

5.5 Frequency Distribution of Node Visiting 111

Chapter 6 -- Conclusions 120

APPENDIX 1. Description of Search Tasks 128

APPENDIX 2. Questionnaire 130

APPENDIX 3. Second-order Transition Probability Matrices 131

REFERENCES .... 13[illegible]

VITA 150

List of Tables

Table	Description	Page
1	State (Zero-order Transition) Probability Matrix	70
2	First-order Transition Probability Matrix	71
3	State (Zero-order Transition) Probabilities for Different Gender Groups	80
4	First-order Transition Probabilities for Different Gender Groups	81
5	State (Zero-order Transition) Probabilities for Different Task Groups	85
6	First-order Transition Probabilities for Different Task Groups	86
7	State (Zero-order Transition) Probabilities for Different Experience Groups	90
8	First-order Transition Probabilities for Different Experience Groups	91
9	State (Zero-order Transition) Probabilities for Different Background Groups	95
10	First-order Transition Probabilities for Different Background Groups	96
11	Result of the Three-way ANOVA	98
12	Mean AB Rates	99
13	Path Length Data	106
14	Grouped Path Length Data	107
15	Data of Frequency of Node Visiting	112

16	Observed vs. Expected Data for the Zipf Distribution	114
17	Comparison of fitting of Different Models	117

List of Figures

Figure	Description	Page
1	Sample Screen 1 of the Experimental System	40
2	Sample Screen 2 of the Experimental System	41
3	Sample Screen 3 of the Experimental System	42
4	Sample Screen 4 of the Experimental System	43
5	Sample Screen 5 of the Experimental System	44
6	Sample Screen 6 of the Experimental System	45
7	Mean AB Rate Values	100
8	Frequency Distribution of Node Visiting	115

List of Appendices

Appendix	Description	Page
1	Description of Search Tasks	128
2	Questionnaire	130
3	Second-order Transition Probability Matrices	131

Chapter 1. Introduction

Hypertext is an associative information management system, which uses a series of links and nodes to associate units of information. (Franklin, 1989) It allows users to access related information through machine-supported links. During the past decade, the development of hypertext technology and graphical user interfaces (GUIs) has brought about significant changes in the construction and use of electronic information systems. Unlike traditional information retrieval systems, in which users access information from a large database using Boolean operations on keyword strings, users of hypertext systems "navigate" through the information database by following links from one piece of information to the next. Such an architecture encourages users to find information by browsing, i.e. following a likely path from one node to another until they reach their objective (Campagnoni & Ehrlich, 1989, p. 212).

Before describing the objectives and design of this study, it is necessary to define some basic, frequently-used terms. Hypertext is a network of nodes connected by links. A node is the basic unit used to store information. For example, a node could be the table of contents or the text of a paper. A link is a relationship between two nodes. For example, the node that contains the citing paper and the node that contains the cited paper can be connected by the link citation. The table of contents node and the text node can have the link title-text. In hypertext systems, users get information by visiting nodes, since information is stored in nodes. Because there are different types of links among nodes, there are different ways of moving from one node to the other. In other words, there are several search options in a hypertext system, such as searching through Table of Contents (moving from table of content node to a specific text node), searching through citation (moving from citing paper to cited paper through the citation link). Different users choose different search options and consequently go through different search states and form different search paths. In this study, a path is defined as a sequence of nodes visited during a search. The path length is defined as the number of nodes in the path. A state is the situation the search is in. To illustrate these concepts, suppose a user who wants to get some general knowledge about hypertext is searching in the Hyperties version of Hypertext on Hypertext (a full text database on hypertext). First, the user searches through Table of Contents, and decides to read the article "Hypertext 87 Workshop". So he clicks on this title and a brief description of the article is displayed. Then he clicks on the title again and the article is brought on to the screen. The article mentioned Ted Nelson, one of the pioneers of hypertext systems. The user wants to know more about Nelson, so he clicks on the highlighted term "Ted Nelson" and Nelson's biography is retrieved. After reading the biography, the user moves back to the article "Hypertext 87 Workshop" to read some more. The paper "Hypertext and Oxford English Dictionary" is cited in the article and the user is curious about this, so he clicks on the title and the paper is displayed on the screen. The user is not very interested in this paper, so he decides to look for something else. But he forgets how he arrived here (he is lost), so he displays the "Path History" and then decides to quit the search. In this example, the user went through the following search states: searching through Table of Contents, displaying title, displaying article, searching through highlighted term, moving back to a node already visited, searching through citation, displaying path history. The path for this search (the nodes visited) is: Table of Content, Hypertext 87 Workshop, Ted Nelson's biography, Hypertext 87 Workshop, Hypertext and Oxford English Dictionary, Path History. The path length is six since there are six nodes in the path. The frequency of node visiting is: one node is visited twice, four nodes are visited once.

Many studies have been done on hypertext systems covering various topics such as: appropriate data models, size of nodes, types of links, user interfaces, users' search strategy, search mechanism, disorientation problems and maintenance of hypertext systems. Studies have been carried out on the search states and paths users take in hypertext systems. For example, Marchionini and Shneiderman's study (1988) showed that when searching for factual information (i.e. the user has a specific search topic), the predominant strategy was to use an index, while in a museum setting where searching involves mainly browsing knowledge, the majority of selections were through the embedded menus. Campagnoni and Ehrlich (1989) conducted a study on a hypertext-based help system and found that the predominant search made was "browsing" (characterized by scanning tables of contents and paging through topics), rather than employing the index ("analytical search"). However, little research has been carried out to study the search states comprehensively, that is, to cover each possible state, not only the browsing or keyword searching state but also all the other states such as displaying search results or getting help.

Numerous mathematical models have been developed for traditional information retrieval systems; and simulation studies have been carried out to study the information retrieval systems based on these mathematical models (Cooper, 1973; Griffiths, 1977; Nelson, 1981). However, very few mathematical models or simulation studies for hypertext information systems have appeared.

Marchionini and Shneiderman (1988, p. 71) pointed out the need for research on hypertext systems: "Since hypertext systems have a brief history of application, we have sparse evidence for their effectiveness, let alone proven principles to guide design". The purpose of the present study is to expand our knowledge of the actual use of hypertext systems. Specifically, I hope to discover the search state and path patterns through which users search for information in hypertext systems. The questions to be investigated in this study are as follows: What is the state use distribution in a hypertext system? What is the probability of going from one state to another state? What states are most likely to follow one another? Do different user groups perform differently in terms of search state patterns? If so, how do they differ? Do different search tasks (general task vs. specific task) influence users' choice of states and if so, how do they influence them? How do users choose search options (analytical searching vs. browsing)? What is the path length distribution? How frequently do users go back to a node already visited? The main method for describing the search state and path patterns in this study is through the development of mathematical models as described below.

The sequences of states searches go through can be described mathematically by transition probability matrices. The Markov chain analysis based on these matrices provides a quantitative technique for describing users' behaviour. The Markov chain model is both descriptive and predictive. The data in each cell of a transition probability matrix is the probability of going from the corresponding row state to the column state. Thus, a transition probability matrix describes a pattern of movements through states that can provide a map of the user behaviour. An analysis of user behaviour in this manner not only describes what proportion of a search is involved in string searching, index searching, etc., but also what states or actions are most likely to follow one another. The comparison of transition probability matrices across different user groups and search tasks revealed the search patterns of corresponding user groups. In addition, a three-way analysis of variance test was conducted to study the effects of gender and search task as well as searching experience on users' search options (analytical searching vs. browsing).

The path length distribution and the frequency distribution of node visiting (how many nodes were visited once, how many twice, etc.) were developed and tested using goodness-of-fit tests. These distribution models can be applied to simulation studies and queuing problems.

The establishment of probabilistic models for hypertext search state and path patterns provides valuable information in evaluating a system's existing operation and in refining future designs. They can help us better understand users' search behaviour and the search processes involved. They also provide a background for examination of systems via simulation studies. The significance of finding the users' search state and path pattern in hypertext systems can also be seen from Marchionini and Shneiderman's statement (1988, p. 74) that: "Although information units retrieved by the system are easily collected for analysis, the analysis of the search process causes more problems. Examination of paths taken and decisions made in jumping to other nodes allow us to make inferences about users' cognitive activity and provide evaluative data on systems effectiveness".

This thesis begins with an introduction to the Markov model, which is the major model used in the study (Chapter 2). A literature review (Chapter 3) provides an overview of related studies. Chapter 4 gives a detailed description of the methodology, including the experimental system, the subjects, the search tasks, data collection, and data processing. Results of the study are reported in Chapter 5 and the conclusions drawn from the results are summarized in Chapter 6.

Chapter 2. Basic Knowledge of Markov Models

Probability is the study of random experiments. A stochastic process is an indexed collection of random variables $\{X_t\}$, where the index t runs through a given set T. Often T is taken to be the set of non-negative integers, and X_t represents a measurable characteristic of interest at time t.

An event B is said to be independent of an event A, if the probability that B will occur is not influenced by whether or not A occurs, (i.e., if P(B/A)=P(B)). If the probability of B changes value with the occurrence or non-occurrence of A, then B is conditionally dependent upon A.

A stochastic process $\{X_t\}$ is said to have the first order Markovian property if $P\{X_{t+1}=j/X_0=k_0, X_1=k_1, \ldots, X_{t-1}=k_{t-1}, X_t=i\} = P\{X_{t+1}=j/X_t=i\}$ for t=0, 1, ... and every sequence, i, j, k_0, k_1, .., k_{t-1}.

This Markovian property can be shown to be equivalent to stating that the conditional probability of any future event, given any past event and the present state $X_t=i$, is independent of the past event and depends only on the present state of the process. The conditional probabilities, $P\{X_{t+1}=j/X_t=i\}$ are called transition probabilities. If, for each i and j,

$$P\{X_{t+1}=j/X_t=i\} = P\{X_1=j/X_0=i\}, \qquad \text{for all } t=0, 1, \ldots,$$

then the (one step) transition probabilities are said to be stationary and are usually denoted by p_{ij}. Thus, having stationary transition probabilities implies that the transition probabilities do not change with time. The existence of stationary (one step) transition probabilities also implies that, for each i,j, and n (n=1, 2, ...),

$$P\{X_{t+n}=j/X_t=i\} = P\{X_n=j/X_0=i\},$$

for all t=0,1, ... These conditional probabilities are usually denoted by $p_{ij}^{(n)}$ and are called n-step transition probabilities. Thus $p_{ij}^{(n)}$ is just the conditional probability that the random variable X, starting in state i, will be in state j after exactly n steps (time units).

Since the $p_{ij}^{(n)}$ are conditional probabilities, they must satisfy the properties,

$$p_{ij}^{(n)} \geq 0, \qquad \text{for all } i \text{ and } j, \text{ and } n=1, 2, \ldots$$

$$\sum_{j=0}^{m} p_{ij}^{(n)} = 1, \qquad \text{for all } i \text{ and } n=1, 2, \ldots$$

A convenient notation for representing the transition probabilities is the matrix form,

$$p^{(n)} = \begin{vmatrix} p_{00}^{(n)} & \cdots & p_{0m}^{(n)} \\ \vdots & \ddots & \vdots \\ p_{m0}^{(n)} & \cdots & p_{mm}^{(n)} \end{vmatrix}, \quad for\ n=1,2,\ldots$$

It is now possible to define a Markov chain. A stochastic process $\{X_t\}$ (t=0, 1, ...) is said to be a finite-state Markov chain if it has the following properties: (Hillier and Lieberman, 1972, p. 404)

(1) a finite number of states,

(2) the Markovian property,

(3) stationary transition probabilities,

(4) a set of initial probabilities, $P\{X_0 = i\}$ for all i.

The process $\{X_n\}$ is a Markov chain of order t if the conditional probability

$$P \{ X_n = a_n / x_m = a_m,\ m < n \}$$

is independent of the values of a_m for $m < n-t$. (Billingsley, 1961, p. 12)

The "order" is the number of previous states that will influence the probability of occurrence of the present state. For example, if the search goes through the following states: searching through the index, searching through citation, and string searching, and if the probability of going through the string searching state is influenced only by the state "searching through citation" but not by the state "searching through the index", then it is a first order Markov chain. If the probability of string searching is influenced not only by the state "searching through citation" but also by the state "searching through the index", then it is a second order Markov chain. If the probability of the string searching is not influenced by any previous state, then it is a zero order Markov chain.

Chapter 3. Literature Review

The related studies are reviewed from the following four aspects:

(1) Application of Markov models in library and information science.

(2) Research in hypertext information retrieval systems.

(3) User Studies in information retrieval systems. This section of the literature review covers the comparisons of the different user groups in this study.

(4) Probabilistic models for traditional information retrieval systems. Literature reviewed here is related to the path distribution models developed in this study.

3.1 Application of Markov Models in Library and Information Science

Many situations exist in which human choice and behaviour can be described quite realistically by using probabilistic models, which allow for conditional behaviour, rather than deterministic models, which do not. The Markov theory in recent decades has found many applications in the physical and social sciences and in engineering and business. It has been recognized as an important technique to characterize many non-deterministic processes (Pao & McCreery, 1986). In library and information science however, although a substantial literature has been built attempting to apply probability and stochastic process theory, Markov models have been used in only three areas: information retrieval systems, library circulation and bibliometric applications.

3.1.1 Markov Model for the Information Retrieval Systems

One of the first applications of a Markov model to the online searching process was conducted by Penniman (1975a, 1975b) (Penniman (1975b) mentioned one earlier study of employing this model to analyze online behaviour, an unpublished Ph.D. dissertation by Stabell C.B.). Starting from the notion that online information retrieval is a form of human-computer communication, Penniman developed a Markov model for online search patterns which posits that a state or event occurs in a predictable fashion based upon the preceding state or event. Penniman reduced a search to a number of user command/system response states, such as index search, logic formation, and document display, which were mutually exclusive and jointly exhaustive of all the possible states of an online search. Using data collected from monitoring over 200 hours of use of Battelle's BASIS online retrieval system, he developed a transition probability matrix. Penniman calculated not only zero order and first order transition probabilities but also higher order transition probabilities. He concluded that "progressively higher order models provided more predictive power. In essence, this means that the more historical data on user behaviour that is available (t-1, t-2, t-3, etc.) the better we can predict the next state of occurrence. Thus the two-dimensional matrix ... is not as powerful as a three-dimensional matrix, but is more powerful than the single-dimensional matrix of a zero-order model." (Penniman, 1975b, p148). This conclusion was used by later people to build high order models (Chapman, 1981). This procedure may be inappropriate in that it may mislead people to blindly add higher orders to models without increasing the predictive power. The order of the search process indicates an upper bound for prediction. If a search process has a second-order nature, which means the past two states influence the probability of present state, third- or higher order models will not provide more predictive power.

In 1981, Chapman published a paper applying the same techniques to study the sequence of commands in online information retrieval. Zero through fourth order Markovian analysis of individual commands and strings of like commands were performed to compare the searching procedures used by three different classes of users. The purpose of this study was to examine the extent to which different groups of users of an online bibliographic retrieval system exhibit different language behaviour. The user groups studied differed in the type and amount of training received in the techniques of online information retrieval. Statistical analyses of searching patterns were carried out to discover the differences in the usage of the query language which may be attributable to differences in user backgrounds. However, no statistical test was conducted to determine the appropriate order of the Markov model. In other words, it is still unknown how many of the previous states influence the choice of the current state. Chapman comparea zero- and first-order transition probability matrices of different user groups and found that "significant difference in lower-order data appear to be sufficient to account for differences in transitions of a higher order" (Chapman, 1981, p. 327). Unfortunately, this statement was later wrongly used as the argument of not developing high order models. It was misunderstood because it does not mean that higher order models do not provide more information than lower order models.

Borgman (1982) used the Markov state transition analysis technique as applied by Penniman to study the Ohio State University online catalog usage patterns. The analysis sought to identify user search patterns, and comparisons were made among campus libraries to determine if any significant difference between user populations could be identified. In 1986, Borgman reported another user study which used the Markov model. Zero- to second-order transaction probability matrices were developed but no test as to which order is appropriate was conducted.

Later, Tolle published several papers reportir.· a Markov chain analysis of OPAC usage patterns (Tolle, 1985, 1984a, 1984b, 1983a, 1983b). The objective of this study was to discover the extent to which system features are used and to determine the actual user patterns when conducting an information search. A number of discoveri were made from the transition probability matrices. For example, because there is a high probability of going from one error state to another error state in the transition probability matrix, it was concluded that there was a definite tendency to repeat errors; i.e., once an error is made, there is a high probability of transition to another error, even at the high order of transitions (Tolle, 1984a, p. 254). Another observation is that the users tend to remain in the type of search that was initiated originally (Tolle, 1984a, p. 255). These discoveries illustrate the potential of transition probability matrices for studying the information retrieval behaviour.

Tolle also calculated high order transition probability matrices based on Markov chain analysis and conditional probabilities. Kolmogorov-Simirnov (K-S) tests were conducted to compare the theoretical versus observed distributions. But the goodness of fit test results were not reported (Tolle, 1985, 1984a, 1984b, 1983a, 1983b). Tolle did not state which order of Markov model was best. Neither did he reject any order of model based on goodness of fit tests. He claimed: "the model used may be rejected by certain individual differences but still be quite useful, especially as a simulation model. The data must be examined and not rejected simply by the nonfit at a given significance level" (Tolle, 1983a, p. 22). In Tolle's opinion, higher order analysis and the cross comparison of systems cannot help but increase our knowledge of online catalog use and in turn improve both existing and future systems (Tolle, 1983b, p. 160). Obviously, Tolle was not interested in finding out which order of the Markov model is appropriate to describe the OPAC usage state pattern. That is, he did not try to find out how many of the previous states influence the probability of going from the present state to the next state. His purpose in conducting goodness of fit tests was not to check if the Markovian properties exist, and in fact it is not clear, from his published papers, why he conducted goodness of fit tests. In my opinion, it is not appropriate to apply the model to simulations without first validating the model. I believe that high order Markov chain analysis incorporating goodness of fit tests can help to discover which order of the Markov model is appropriate to describe the state pattern.

One direction for future research, identified by Tolle, is the following: the co-occurrence of states must also be determined and be input into the framework that the models describe. These associations may allow quantification by geometric or spatial representation using multidimensional scaling techniques which can describe the similarity of states (Tolle, 1984a p. 256). When concluding this research, Tolle wrote: "The performance measurement (monitoring) and evaluation (analysis) provide valuable information and techniques to evaluate a system's existing operation and to attempt to refine future online system design. It also allows the development of models which may represent the user search behaviour and thus provide a method for examination via simulation" (Tolle, 1984b, p. 129). He considered simulation as one of the future avenues of research (Tolle, 1983a, p. 113). These are interesting research directions worth further exploration. However, an online search on INSPEC, LISA(Library and Information Science Abstracts), ERIC and ISA (Information Science Abstracts) databases in DIALOG did not reveal any subsequent papers by Tolle in this area.

Harris (1986) applied transition analysis techniques similar to those of Penniman to larger segments of the online search process (here, a segment of search comprises all commands leading up to and including a DISPLAY, TYPE, or PRINT command or commands). Searches from two sources (student vs. professional searchers) were compared in order to determine whether different patterns of behaviour would be found with different types of users (experienced vs. inexperienced) and different search requests (several different search requests vs. single search request). Unfortunately, there is a major mistake in the design of the study which invalidates the result. The twenty-four searches from the student source were conducted by twenty-four different individuals while the twenty-five searches from the professional searchers were conducted by two different people. When conducting the statistical test to compare the two types of searches, the assumption is that the searches were independent of each other. Obviously, the searches conducted by the same professional searchers were not independent of each other, but correlated in the way that they are conducted by the same person. Two searchers are definitely not an appropriate sample to represent the professional searchers as a whole.

Recently, Marchionini (1989) conducted a study of elementary school children searching a full-text electronic encyclopedia on CD-ROM. First-order transition probability matrices for different user groups were developed and compared, revealing that younger searchers generally favoured query refining moves and older searchers favoured examination of the title and text moves.

When reviewing research in online searching, Foster (1986) commented on the application of stochastic analysis (Markov chain analysis in particular) to online information seeking behaviour. She considers this to be a solid methodology for search pattern analysis using information that the computer has made available. Data provided by online monitoring presents the first real chance to study information seeking behaviour directly, and stochastic analysis is a significant methodological contribution to that study. Penniman (1975b) summarised the usefulness of the stochastic modelling of user behaviour in studying online searching as follows:

(1) it provides a quantitative technique for describing behaviour;
(2) it provides a statistical rigour to the comparison of behaviour of various sub-groups of individuals or comparison of sub-samples;
(3) it provides a predictive as well as descriptive model.

Foster agrees with the first two points but considers the third point as limited to some extent since the stochastic model considers the form of a search (its command structure), and ignores both the content and desired outcome. Significant variables such as the choice and order of the search vocabulary or size of document set retrieved are not considered as variables which may affect the progress of a search.

3.1.2. Markov Model in Library Circulation Process

Philip M. Morse's 1968 work, Library Effectiveness: A Systems Approach, has had the greatest influence on Markov model applications to the library circulation process. Morse used a least-square formula for finding the mathematical relationship of the mean circulation of a group of books, all circulating a given number of times in year t+1 as a function of circulation m for previous year t, where m is a given number of circulations. To obtain his transition probabilities, he assumed a Poisson distribution to predict the chance of getting circulation n in year t+1. However, no goodness of fit test was conducted to validate the model; and no higher order of transition probability matrix was developed.

Ching-chih Chen's book (1976) "Applications of Operations Research Models to Libraries; A Case Study of the Use of Monographs in the Francis A. Countway Library of Medicine, Harvard University," based on her doctoral dissertation, reports efforts to test circulation models developed by Philip Morse. In the preface of her book, Chen states; "The model developed by Morse was tested by using quite limited experimental data obtained at the MIT Science Library in the early 1960s. In order to test the validity and accuracy of his probabilistic models, further research in a different library environment was necessary, and more up-to-date and comprehensive data were needed." The database Chen used consisted of approximately 12,000 circulation records collected in four selected complete months of 1973 in the Francis A. Countway Library of Medicine, Harvard University. As Bookstein (1977) commented, though Chen intends to show the validity of the Morse Markov model with Poisson transition probabilities, the Markov nature of the model is never explicitly tested. Similarly, although the Poisson distribution is shown to "correspond quite well with the experimental data", no statistical tests of goodness of fit are given.

Coady (1981), in his doctoral dissertation "The Applicability of Markov Models to the Circulation of Social-Science Monographs in a Large Academic Library," applied

Morse's Markov model to social science monographs. The goal of Coady's study was to check circulation data for Markovity and therefore submit evidence toward the validation or invalidation of the use of the Markov process in the analysis of the use of social science monographs. He used circulation data from Ohio State University Libraries during the period 1975-1978. The Chi-square test for Markovity showed that all fourteen test groups were without Markovity.

Utilizing eleven years of circulation transaction data from a university library, Beheshti and Tague (1984) tested the reliability of the basic assumptions underlying the Markov model's parameters. It was found that, contrary to Morse's assumptions, one of the model's two parameters is time dependent.

Later, Tague and Ajiferuke (1987) applied the Poisson model to a large eleven-year university circulation data set. The goodness of fit tests indicate that the model does not fit the data. The set of non-circulated items is larger than that predicted by the model.

3.1.3. Bibliometric Application of Markov Model

In 1971, a Markov chain model of science development was proposed and developed by Zunde and Slamecka. It was based on the assumption that scientific development is incited and sustained by "stimulators", the most significant of which are communications containing the latest information on scientific inquiry and research. A significant proportion of these stimulators is reflected in citation data. So a Markov model was applied to a sample of citation data in social science. The row and column vectors are scientific disciplines. The probability that a publication in discipline A cites a publication of discipline B was used as the transition probability in the Markov model. The analysis of this model shows a tendency of shifting emphasis in scientific enquiry among different disciplines.

Also in 1971, Goffman used the theory of Markov chains to describe the movement of authors among subtopics of symbolic logic based on their publication data.

Similarly, Pao and McCreery (1986) applied a Markov chain model to authors' publication data to describe their movements among sub-areas of a discipline. They also suggested other useful areas of bibliometric applications using the Markov chain model, such as tracking the trend of publishing among a group of journals by the shift of citations among these journals. Although conducted 15 years later, Pao and McCreery's study provided no further developments in methodology beyond those of Goffman.

Thus, there has been little research carried out in the application of Markov models in the bibliometrics area, and the scale of these studies is small compared to those dealing with library circulation and information retrieval systems.

3.2 Research in Hypertext Information Retrieval Systems

Hypertext systems are based on the concept that there is a similarity between human associative memory and a set of text fragments that may be linked together in any order. Through different links, hypertext systems allow users to traverse complex networks of information quickly. Marchionini (1987) discussed the importance of browsing for hypertext systems. McAleese (1989) considered browsing to be central to effective hypertext systems and examined different browsing strategies such as scanning and wandering. However, he did not provide any way to classify use patterns into these different types of browsing, so they cannot be applied to label path patterns.

Browsing through nonlinear networks of frames often leaves people with a general feeling of disorientation, being lost, or of losing context. This is the hypertext system users' navigation problem: knowing where one is, where one wants to go and how to get there from the present location.

Much research has been done on the disorientation problem. Generally, there are two approaches to this problem: the system design approach and the cognitive science approach. In the system design approach, there are two main methods: creating orientation tools and providing a pre-defined path or PATH. A path is an ordered traversal of some of the links in a hypertext system.

As early as 1982, Mantel studied the disorientation problem in her Ph.D. thesis: "Disorientation Behaviour in Person-Computer Interaction". She conducted a user experiment with the ZOG system to study the hypothesis that disorientation is a failure of the individual to adequately acquire and organize information about the environment for effective navigation, and the individual's recognition of this failure.

Some hypertext systems such as HyperCard provide maps to help users orient themselves. Other systems such as Hyperties provide a list of nodes the user has traversed. In Nielsen's system (1990), the time since the node was visited is also provided in the "History List". Many systems provide an orientation tool called a "browser", which is a node containing a structural diagram of the network of nodes. Foss (1989) describes four types of browsers -- graphical history list, history tree, summary boxes and summary tree -- which have been implemented in the Xerox NoteCards hypertext system. These four browsers support the processes of pursuing and returning from digressions and facilitate the integration of the diverse sets of materials that have been browsed. Bernstein (1988) described an array of orientation tools, including book marks, thumb tabs etc., which help users to choose a course through a complex information network without undue confusion or discomfort.

The concept of a PATH, or pre-defined ordered traversal of some links in a hypertext, has been a part of the hypertext notion from its early formation. (In order to distinguish the specific PATH referred to here from the usual path mentioned elsewhere, the specific PATH will be written in uppercase, PATH). Although a PATH can help to solve two major problems with hypertext systems, namely user disorientation and high cognitive overhead for users, its value has not been recognized. Zellweger (1989) constructed a framework for understanding hypertext PATH mechanisms and explores the basic issues surrounding them. Given this framework, Zellweger reviewed PATH mechanisms that have been provided in other hypertext systems. The author also described the Scripted Documents system, which has been developed to test the potential of one powerful PATH mechanism.

PATHs can help to solve the disorientation problem because users are less likely to feel disoriented or lost when they are following a pre-defined PATH rather than browsing freely. When designing a PATH, hypertext designers must determine an appropriate order of presentation. They need the information about the route users usually take in hypertext systems. The transition probability matrix which will be developed in this study describes path patterns users will take when they search freely in hypertext systems. Obviously, such information will be very useful for designing a pre-defined PATH.

Hendry (1989) investigated the relationship between comprehension and navigation in a hypertext environment and found that there was little relationship. He realized the need of a deeper investigation of the indices for characterizing navigation. In discussing browsing and navigation issues, Dumais (1988) pointed out that "there is little known about the range of materials or tasks for which network navigation is an appropriate searching technique." Barfield et al. (1990) listed specifically the following questions on which much research remains to be done: Which types of tasks are more suited for hypertext? Will novice and experienced computer users perform differently with hypertext systems? If so, what are the implications for training?

Nielsen (1989) compared 92 published benchmark measurements of various usability issues related to hypertext and found the two most important issues to be: (1) individual differences among users (different people will perform very differently); and (2) the effect of different tasks (people with different tasks will use hypertext systems in different ways). These two factors will also be the main factors to be investigated in this study as stated in Chapter 5.

Although there is no standardized definition of analytical searching and browsing, the distinction between the two types of search patterns made by Marchionini and Shneiderman (1988) is fairly clear. They said: "As distinguished from analytical searching, which requires formation of specific, well structured queries, browsing involves the generation of broad query terms or other indicators of relevance such as citations, authors etc., and scanning much larger sets of information in a relatively unstructured fashion. Such search features as Boolean connectives, string searching, proximity limits, scope limits, and truncation facilitate rapid access to information, but cause additional cognitive load on the part of the user and substantial processing of the database itself. Browsing features on the other hand, allow casual, low cognitive load exploration, but are typically inefficient for directed search tasks or fact retrieval."

As Marchionini and Shneiderman (1988) indicated, determining criteria for optimal mixes of browsing and analytical support is critical to development. Balancing retrieval power with ease of understanding is a central problem for designers of future systems. It was for these purposes that this study also investigated users' search options in terms of analytical searching vs. browsing.

In a study conducted on information retrieval using a hypertext-based help system (Campagnoni & Ehrlich, 1989), it was found that analytical searching (in that system, use of indexes) was most frequent in subjects who were members of the computer system support staff as compared to those who did not have a lot of computer experience. It was found that the tendency to use the indexes was related to the type of question. Users were most likely to resort to the analytical strategy when the question was stated in such a way that the table of contents would not be parcicularly helpful in locating the answer.

Marchionini and Shneiderman (1988) reported an experiment on

Hyperties. When subjects were asked to perform efficiently in searching for specific factual information in a Hyperties database, the predominant strategy (14 of 16 subjects) was to use the alphabetical index (analytical search). In contrast, a log of usage in two museums over eight months showed that more than two-thirds of all selections were through the embedded menus, thereby demonstrating the orientation toward browsing in a museum setting.

Gary and Shasha (1989) conducted an experiment to compare the search effectiveness of two versions of the same database -- one with hypertext links and one without. They concluded that in information systems, undirected explorations without a specific question in mind are different from direct searches for a specific piece of information. Links may be more useful when searches are undirected. A person who is generally interested in a topic might follow links that pique curiosity, without worrying too much about getting lost. But faced with a specific question and a set of properties and values to which the query can be mapped reasonably well, the user may be better off not using the links.

3.3 User Studies in Information Retrieval Systems

User studies of information retrieval systems often control for the amount of prior experience subjects have with computerized information retrieval and subjects' other individual differences.

Vigil (1988) argued that the more experience a searcher has, the larger would be his or her problem-solving domain - thus the greater the capacity to look at a complex series of search statements and determine what has been done and what needs to be done next.

Penniman (1975b) found significant difference between novice and the experienced users by comparing the first-order Markov models. Fenichel's research (1981) shows that novices searched significantly more slowly than experienced searchers; their searches were longer in total online time and they entered fewer commands per minute. Also, the novices made about twice as many errors as the experienced subjects. Marchionini et al. (1990) reported a study in a HyperCard database comparing the effects of search expertise and subject expertise. Both groups of experts outperformed a novice control group.

In an empirical study of a Boolean logic based information retrieval system, Borgman (1986) found a relationship between academic orientation (academic discipline chosen by the subjects) and information retrieval performance. Subjects with science or engineering majors were more likely to pass the benchmark and subjects with social science or humanities majors were more likely to fail. In discussing the implications of this relationship, she stated: "most IR studies have been done on the commonly available Boolean (command driven) retrieval systems and we would expect that these interfaces load heavily on reasoning ability. Hence, we might find that command driven interfaces are best for technical populations, while some other interface might be superior for a nontechnical population" (Borgman, 1989, p. 248).

In the same study mentioned above, Borgman found a significant difference between sexes in search-state transitions. On simple tasks, men and women reflected different patterns of system use at all three levels of zero-, first-, and second-order transitions ($P<0.01$ for zero and first order. $P<0.001$ for second order). On complex tasks, men and women also reflected different patterns of use at all three levels, although less strongly ($P<0.01$ for zero and second order. $P<0.05$ for first order). No detail as to how men and women differed was reported.

It will be interesting and useful to find out if these performance differences for different user groups in traditional information retrieval system also characterize use of hypertext information retrieval systems.

3.4 Probabilistic Models for Traditional Information Retrieval Systems

Although few mathematical models have been developed for hypertext information retrieval systems, there are numerous mathematical models for traditional information retrieval systems. Among the many types of probabilistic models of the information retrieval systems, only models for the distribution of index terms will be reviewed here, since they correspond to the path distribution models which will be developed in this thesis.

Since distribution models can describe the index terms with a small number of parameters, they can be used to characterize a particular system. Distributions are also very useful in simulation for the same reasons. There are two types of distributions which are commonly used to summarize frequency data: the size frequency and the rank frequency. In size frequency distributions, the distribution function $g(x)$ is the number of terms which have x postings, x=1, 2, 3, ..., X_{max}. G(x) is the cumulative distribution of g(x). In rank frequency distributions, the distribution function f(r) is the number of postings of the r-th ranked terms, r=1, 2, 3, ..., R_{max}, where the terms are ranked in descending order by the number of postings. F(r) is the cumulative distribution of f(r).

There have been many attempts to describe the distribution of the occurrence of words in English language text. Perhaps the most quoted is that of Zipf (1949), who suggested that f(r)=c/r for a large number of different samples of text, where r is the frequency rank of a word (i.e., r=1 for the most frequent word), f(r) is the number of occurrences of the r-th ranked word, and c is a constant. Later researchers have given many versions and explanations of Zipf's law. For example, Mandelbrot (1953) generalized this function to get a better fit in the form $f(r)=c/(r+d)^b$, where b, c and d are parameters. Many studies of the distributions of index terms have been influenced by Zipf's distribution.

Many empirical studies have been carried out to test the distribution of index terms. One of the earliest studies which collected statistics for the distribution of index terms was by Schultz, Schwartz and Steinberg in 1961. They plotted the distribution of indexing exhaustivity (number of terms used to index a document) and then described the general shapes of the distributions. However, they did not fit mathematical distributions to these graphs.

Later, other investigators tried to fit mathematical models to the graphs. However, goodness of fit tests were not carried out in many cases. Griffiths (1975, 1977) looked at the distribution of index terms for incorporation into a stochastic model for use in simulation. She looked at the distribution of terms from several systems and claimed they follow a generalized Bradford distribution $F(r)=k*\log[(a+r)/a]$ where $k=1/\log[(a+N)/a]$. However, no goodness of fit test was conducted to see how well the data fitted the model.

Recent studies tend to be more rigorous in their mathematical approach and goodness of fit tests have been carried out to help identify models. Nelson and Tague (1985) proposed a split model, which used both a rank function for the high frequency terms and a size function for the low frequency terms, with the point of transition being determined either empirically or by rule. This model was fitted to the marginal empirical term distributions for four document data sets.

Nelson (1981), in his doctoral dissertation, developed a probabilistic model which includes the distribution of terms over documents and over queries, the distribution of exhaustivity over documents and over queries, the distribution of co-occurrences (occurrences of pairs of terms), the distribution of relevant and non-relevant documents over the number of terms matching the query. These distributions were then incorporated into a simulation program using a probabilistic model of term occurrences and co-occurrences.

Chapter 4. Methodology

The questions investigated in this study are: What is the state use distribution in a hypertext system? What is the probability of going from one state to another state? What states are most likely to follow one another? Do different user groups (gender or academic background or search experience) perform differently in terms of search state patterns? If so, how do they differ? Do different search tasks (finding facts vs. browsing knowledge) influence users' choice of states and if so, how do they influence them? How do users choose search options (analytical searching vs. browsing)? What is the path length distribution? How frequently do users go back to a node already visited?

The use of a hypertext information retrieval system by a group of subjects was monitored to collect search state and path data. These data were then used to study the above questions in the following way. Transition probability matrices were developed which describe in matrix form the probability of going from one state to another state. Search patterns of different user groups were then studied by examining the corresponding transition probability matrices. In addition, a three-way analysis of variance test was conducted to study users' search options (analytical searching vs. browsing). The path length distribution model and the frequency distribution of node visiting (how many nodes were visited once, how many twice, etc.) were developed and tested using goodness of fit tests.

4.1 The Experimental System

The experimental system used in this study was the Hyperties version (IBM-PC) of Hypertext on Hypertext. It contains eight full-text papers from a special issue of Communications of the ACM and a few other papers about Hyperties and ACM. The eight papers are research papers on hypertext. All these papers were cut into 307 small text passages or hypertext nodes as retrieval units. The length of a node varied from one to approximately twenty screens. (See Figures 1-6 for a typical Hyperties session)

FIGURE 1. Sample Screen 1 of the Experimental System

PAGE 1 OF 3

TABLE OF CONTENTS

Communications of the ACM

Special Section on Hypertext Systems : Ben Shneiderman

Hypertext 87 Workshop : Ben Shneiderman

Hyperties : An Overview

ACM Press Database and Electronic Products

Table of Topics

Table of Figures

(more . . .)

HYPERTIES - Hypertext system developed at the University of Maryland; the system you are using to read this.

SEE ARTICLE: "HYPERTIES"

NEXT PAGE   RETURN TO "FRONT MATTER"   EXTRA

FIGURE 2. Sample Screen 2 of the Experimental System

```
TABLE OF CONTENTS   INDEX   HISTORY   SEARCH
                                                 PAGE 9 OF 23
ALPHABETIC INDEX (307 articles)
The introductory article is: INTRODUCTION
   NIELSEN 3 : IS TEXT DATA?
   NIELSEN 4 : THE FATHER OF HYPERTEXT
   NIELSEN 5 : PIONEERS HAVE ARROWS IN THEIR BACKS
   PRODUCING THE HYPERTIES DATABASE
   RAYMOND & TOMPA 1 : TITLE PAGE (OXFORD ENGLISH DICTIONARY)
   RAYMOND & TOMPA 2 : THE OXFORD ENGLISH DICTIONARY
   RAYMOND & TOMPA 3 : WHY HYPERTEXT FOR OED?
   RAYMOND & TOMPA 4 : CONVERTING TEXT TO HYPERTEXT
   RAYMOND & TOMPA 5 : CHARACTERISTICS OF THE OED
   RAYMOND & TOMPA 6 : RELATED PROBLEMS
   RAYMOND & TOMPA 7 : CONCLUSIONS
   RAYMOND & TOMPA 8 : REFERENCES
   RAYMOND & TOMPA : FIGURE 1

   --- TURN TO:--A-B-C-E-F-G-H-I-L-M-N-P-R-S-T-V-W-Y-[-_-------------

BACK PAGE   NEXT PAGE   RETURN TO "FRONT MATTER"          QUIT
```

FIGURE 3. Sample Screen 3 of the Experimental System

VAN DAM 3 : TRIBUTE TO TED NELSON — PAGE 1 OF 4

Previous Section: Tribute to Doug Engelbart
Beginning of this paper: van Dam: Title Page

So much for my paean of praise to Doug E · ·lbart. I think we are all here because of him and also because of **Ted Nelson**, the second trailblazer, who coined the word "hypertext" and dozens of other words -- being wordsmith and master showman par excellence and also a polemicist of the first rank. Ted is a self-proclaimed visionary who deserves the title, and he turned on generations of people with Computer Lib/Dream Machines, a landmark work that still today -- I reread a lot of it just a few days ago to prepare myself for this talk -- is good reading. Ted coined that wonderful phrase, "If computers are the wave of the future, displays are the surfboards." Well, I've used that bon mot ever since, and I think he is absolutely right:

NEXT PAGE — **RETURN TO "HYPERTIES"** — EXTRA

FIGURE 4. Sample Screen 4 of the Experimental System

FRISSE 6 : ENHANCING HYPERTEXT RETRIEVAL VIA USER FEEDBACK PAGE 9 OF 9

allude to this metaphor when they describe the interaction between an electronic encyclopedia and a reader as a "conversation with a guide or tutor who accompanies us during our learning adventure" **[WEYE85]** . Implementing programs that truly guide users is one of the most challenging research topics in hypertext research.

Next Section: Implementing Hypertext in Medical Settings

BACK PAGE RETURN TO "HYPERTEXT 87 WORKSHOP" EXTRA

FIGURE 5. Sample Screen 5 of the Experimental System

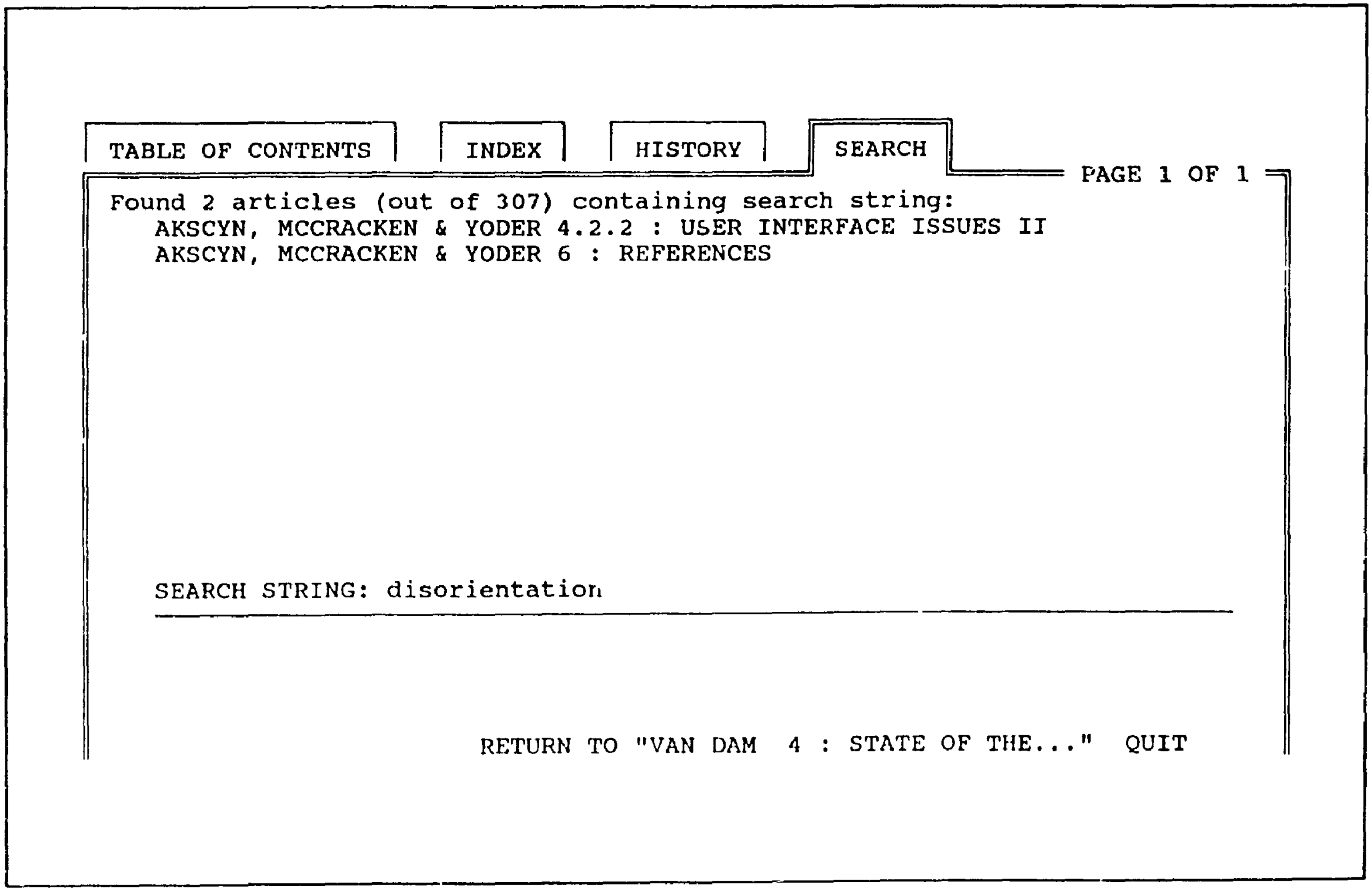
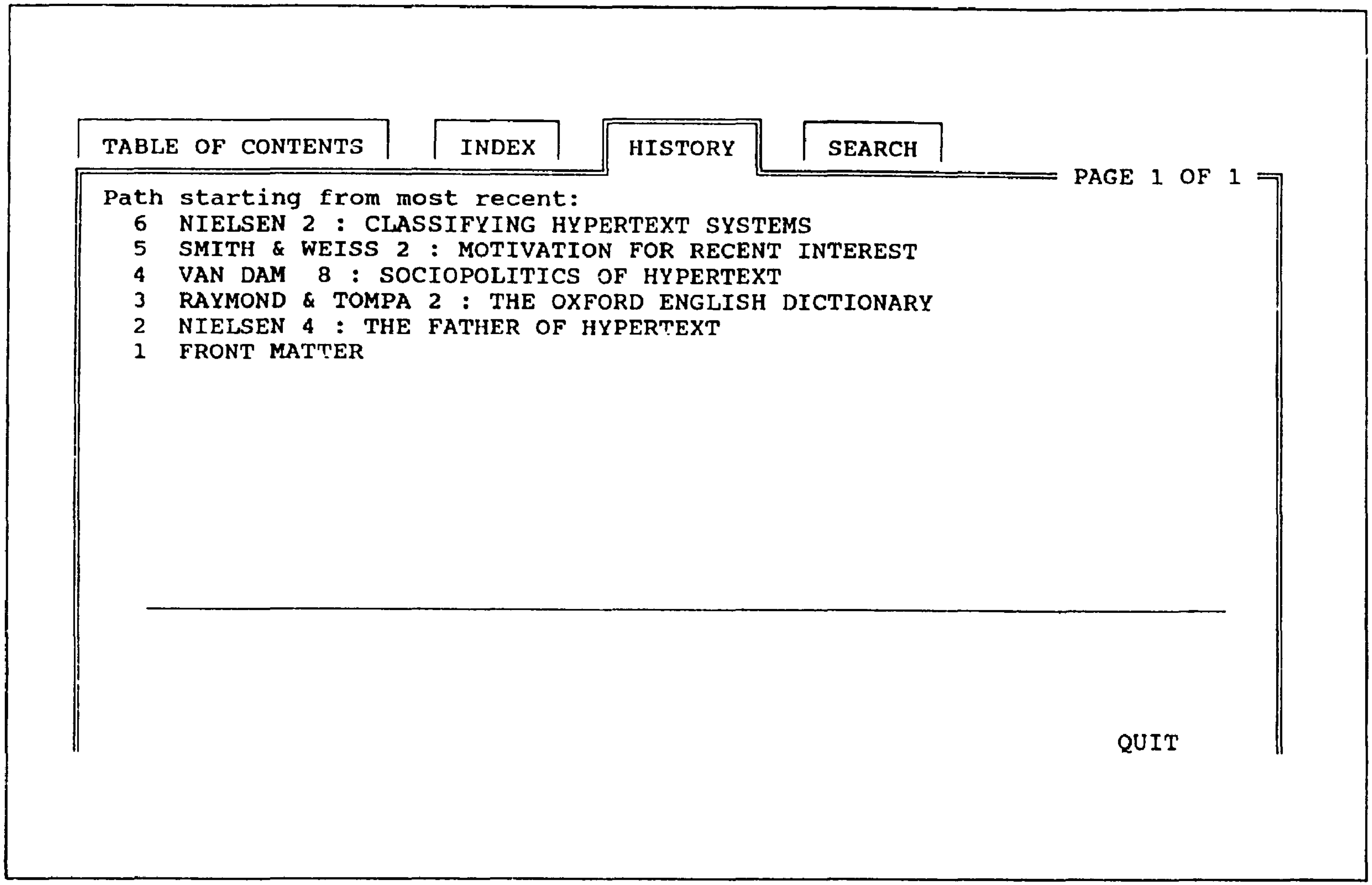


FIGURE 6. Sample Screen 6 of the Experimental System



Hyperties enables users to easily traverse a database of articles and pictures by merely moving the cursor to the title in the table of contents or the index or highlighted terms in the text and pressing the carriage return. A brief description of the content of the chosen node appears at the bottom of the screen. In Figure 1, for example, the title "Hyperties: An Overview" was chosen, two lines of description appears at the bottom part of the screen (in bold). Users may then press another carriage return to bring the chosen node to the screer or move the cursor away to search for other nodes.

Hyperties supports the following five approaches for searching: searching through the table of contents, the index, a citation, or a highlighted term in the text, and string searching. The table of contents is a list of papers in the database. Figure 1 is one the three screens of the table of contents. A single paper was cut into several parts and each part was stored in a node. Each node has a title assigned by the system designer based on the content of the node. The index is a list of node titles arranged alphabetically according to the author(s)' last name. Figure 2 is one of the twenty-three screens of the index. Figure 3 is a typical screen of a text node. The term "Ted Nelson" is highlighted and selectable by moving the cursor to the term and pressing the carriage return. The square bracket "[WEYE85]" in Figure 4 indicates the citation of Weyer's paper which was published in 1985. The full bibliography can be brought to the screen by moving the cursor to the square bracket. String searching enables users to search through an author or a keyword (any word in texts is searchable). The user can type in a single word or two words, but not more than two words. Search terms are automatically truncated on the right, but Boolean connectives are not fully supported. The search result is a list of nodes that contain the search term. Figure 5 shows the result of string search on the term "disorientation". The user can move the cursor to any of the retrieved articles and bring it to the screen by pressing the carriage return.

A single paper is cut into several parts and each part was stored in a node. Nodes that belong to the same paper are connected by hypertext links. Users can move backward or forward in a paper through these links. Figure 3 and Figure 4 illustrate this feature. Moving the cursor to the second line in Figure 3 "Previous Section: Tribute to Doug Engelbart" and pressing the carriage return will lead the user to the previous part of the paper "Tribute to Doug Engelbart". The same can be done to go to the beginning part of the paper through the third line in Figure 3 "Beginning of this paper: van Dam: Title Page". Likewise, the second last line in Figure 4 "Next Section: Implementing Hypertext in Medical Settings" leads to the next part of the paper.

As users traverse the database, Hyperties traces the path and allows them to return to previous nodes. There are two ways of going back to a previously visited node: moving back to the immediate past node and jumping to any remotely past node. Figure 3 illustrates the former option. Moving the cursor to the bottom line RETURN TO "HYPERTIES" will lead to the immediate past node "HYPERTIES". The bottom line of the screen always provides this opt' n. Using this option continuously, users can trace back to any node already visited, all the way back to the first node. The latter option is realized by displaying the path history. The path history node contains a list of nodes already visited during the search. Figure 6 is an example. Users can go back to any of the listed nodes by moving the cursor there and pressing the carriage return.

The limitation of Hyperties is that, unlike Guide (another commercial hypertext software) which has a rich set of link types, Hyperties supports only a single type of link that points from a text string within an article to another article (Glushko, 1990). Another limitation of Hyperties is that the path history which is used to help users in the disorientation situation has a linear form: it is simply a list of the nodes visited. The network structure of nodes and links is not presented in a map form as some other systems do. Therefore, it is difficult for users to perceive the location of the path from a global point of view.

4.2 Subjects

Sixty-one subjects participated in the experiment. Unfortunately, six subjects' search logs were lost or incomplete because of a software or hardware problem. Subjects were drawn from the following two sources: (1) volunteers (twelve subjects), and (2) forty-nine students enroled in the Online Information Systems and Services course at the School of Library and Information Science, University of Western Ontario during the terms summer 1990 and winter 1991. Students were required to conduct the search on the experimental system as an assignment, but permission for recording the search process was given voluntarily. As a result, every student gave permission for the search process to be monitored. Volunteers were motivated by interest in learning a new system and friendship with the author. Students from the online course were motivated by interest in learning a new system and doing a good assignment. The assignment was to conduct a search on the system and report the search result in writing. The assignments were graded by the professors who taught the course during those two terms. Volunteers simply handed the printout of the search result to the author, with no grades being assigned. It was realized that the difference in motivation between volunteers and students doing the assignment might influence the search performance. However, the requirement of a large sample size to develop the models and the time limitation for finishing the research precluded obtaining a more homogeneous subject group. Due to difficulty in obtaining subjects and ethical considerations, the author was able to impose little control on subjects.

A lecture was given by the author to students in the online course on the subject of hypertext systems. During the lecture, the experimental system was demonstrated (including both the browsing function and the keyword searching function), and students were taught how to use it. For volunteers, the same training was provided by the author on a one to one basis. Ideally, the volunteers should have been trained in a group environment, as the students were. Unfortunately, the difficulty of getting the volunteers together made this design impossible.

There was only one computer available for the subjects to conduct the search. A sign up sheet was posted for the subjects to book the time slot(s) they needed to conduct the search. One time slot lasted for half an hour and most subjects booked two time slots. When the booked time was up and the search was not completed, the subject could continue if the next time slot was free. If the next time slot was not free, the subject could negotiate with the coming subject on time arrangement if he/she wished. Some subjects chose to come back later to continue the search. Every subject had an opportunity to complete the search to their satisfaction. The search lasted for about an hour on average (This average is just the author's estimation, for there was no record as to the time duration). Subjects were allowed to continue searching, instead of being stopped at a given time as in some other experiments, because this approach represents the real search environment better. In most real search environments, it is the user, not somebody else, who decides the length of the search. Before conducting the formal search, every subject had an opportunity to try out the system with the author's help until they felt confident to use the system alone. The monitoring of the search process started when the subject began to search alone.

Very few subjects in this study had used hypertext information retrieval systems before the experiment (i.e., they were mainly novice users). This situation imposed a limitation on the study: all the results may be applicable only to novice users since expert users may behave differently and therefore have a different search pattern. But this limitation will not undermine the significance of the research. As the first attempt in developing probabilistic models for path patterns of hypertext systems, the methodology developed in this study can be used in subsequent research with other user groups. Furthermore, with "end-user" searching becoming more and more popular, most users will be novice or casual users.

4.3 Types of Search Tasks

There were two types of search tasks in this experiment. Each subject was assigned one of the two tasks. Type 1 is a general search task while type 2 a specific search task. In the type 1 task, subjects were asked to conduct a search on the experimental system to find out general information about hypertext systems. This information included definition, history etc. for writing up a one-page encyclopedia entry. In the type 2 task, subjects were asked to conduct a search on the experimental system to find out answers to specific questions about hypertext systems, such as "What size should a node be?", and "What type of links should there be?". There were 13 questions listed under the type 2 task. Each subject doing the type 2 task conducted a search to find out answers to one of the questions listed. See Appendix 1 for a description of the two types of task.

Subjects from the online course in the summer, 1990, term conducted the general task as an assignment, while students in the winter, 1991, term conducted the specific task as an assignment. Volunteers conducted the specific task, since there were only thirteen students in the winter 1991 term. Each volunteer freely chose one of thirteen questions for the search. Students doing the specific task were assigned the questions by drawing slips of paper containing question numbers. Since there were thirteen students and thirteen questions, each question was assigned to one student. As a result, the thirteen questions were fairly evenly distributed among subjects conducting the specific task.

4.4 Data Collection

Transaction logs of searches performed on the system were maintained and used as the basic data. A commercial software utility called Total Recall (Brueggeman, 1989) was installed on the computer to collect the data automatically.

Total Recall recorded the whole search process including screen displays and user inputs. The search process was then played back for analysis. One advantage of computer-monitored data, indicated by Rice and Borgman (1983), is that it is unobtrusive and accurate, and thus improves the validity of the study. Subjects were notified that this monitoring was occurring.

The following information about the subject was collected for each search, in addition to the transaction log: the subject's gender, the subject's academic background (social science or humanities or engineering or science), the subject's experience with computerized information retrieval (ranging from 1, no experience to 5, expert), and the search topic. Subjects were asked to fill out a short questionnaire on these questions (see Appendix 2 for the questionnaire). Subjects' opinions about the system were casually solicited after the search.

In order to have sufficient data points to conduct the appropriate statistical tests, the social science and humanities categories were combined to form a non-technical group of users. Likewise, the science and engineering categories were combined to form a technical group of users. For the same reason, search experience was grouped into two types: low experience (rating from 1 to 2), and high experience (rating from 3 to 5). Among the fifty-five subjects whose search log was complete, twenty were males and thirty-five were females. Forty-eight subjects had social science or humanities background while seven subjects had science or engineering background. Thirty-three subjects rated themselves as having low experience and twenty-two high experience. Thirty subjects performed general task and twenty-five subjects performed specific task.

Some subjects conducted more than one search since they were not satisfied with the result of their first search to do the assignment for the course. All these searches were used to develop the Markov models, but only the first searches were used to conduct the analysis of variance test or develop the distribution models since these tests require that there is no correlation among data points. Searches conducted by the same person can be correlated by this person's search behaviour.

4.5 Data Processing

Once the search processes were recorded, they were played back and parsed into different search states to develop the Markov model. The number of nodes visited in each search was determined to develop path length distribution model and the node visit distribution. All the statistical tests in this study were conducted at the significance level of 0.05.

4.5.1 Distribution Models

A size frequency approach was taken in developing the distribution models. Data were plotted to determine the general shape of the distribution, then the models which have a similar curve were fitted. Goodness of fit tests were conducted to identify the model which best describes the distribution.

4.5.2 The Markov Model

Each search was parsed into the following mutually exclusive and jointly exhaustive states:

(1) SEARCHING THROUGH THE INDEX (A paper is stored in several nodes and each node has a title. The index is a list of node titles);

(2) SEARCHING THROUGH TABLE OF CONTENTS (a list of paper titles);

(3) STRING SEARCH (author, keywords - any word in the text);

(4) DISPLAYING THE TITLE OF THE SEARCH RESULT (with a brief description of the content of the article);

(5) DISPLAYING THE RETRIEVED ARTICLE;

(6) PRINTING;

(7) NODE REVISITING (going back to a node already visited);

(8) SEARCHING THROUGH CITATIONS;

(9) SEARCHING THROUGH HIGHLIGHTED TERMS IN THE TEXT;

(10) LINEAR BROWSING (move to a node which contains another part of the paper currently .ng viewed, including previous part, next part, and other non-consecutive part);

(11) DISPLAYING PATH HISTORY; and

(12) USING HELP SCREEN.

The parsing of all the search processes resulted in 4310 transitions. This is not a large enough sample to estimate a twelve-state transition probability matrix if the Markov order is greater than one. Therefore, it was decided that some states should be collapsed to form a coarser-grained matrix.

The first three states (SEARCHING THROUGH THE INDEX, SEARCHING THROUGH TABLE OF CONTENTS, and STRING SEARCH) were combined to form a single state -- QUERYING. SEARCHING THROUGH CITATION and SEARCHING THROUGH HIGHLIGHTED TERMS IN

THE TEXT have the same search nature -- non-linear browsing -- so these two states were combined to form a new state -- NON-LINEAR BROWSING. Since DISPLAY PATH HISTORY and USING HELP SCREEN have the same purpose -- getting help, they were combined into a single state -- GETTING HELP. As a result, the following eight states are listed in the transition probability matrix: QUERYING, DISPLAYING TITLE, DISPLAYING ARTICLE, PRINTING, NODE REVISITING, NON-LINEAR BROWSING, LINEAR BROWSING, and GETTING HELP.

As is indicated in the Literature Review, the development of a Markov model involves determining the appropriate order first and then using the estimated transition probabilities to describe the search process and compare the search patterns of different user groups. However, there is a limitation on the number of orders in calculating the transition probability matrix, since the possible combinations increase geometrically with states and orders. Therefore, the maximum order to be tested had to be decided first. A log-linear model was then used to test the order of Markov model. Bishop et al (1975, p. 269) gave a log-linear model to test the first- and second-order Markov chain. The model developed in this study (presented below) is an extension of Bishop et al's model and can be used to test any order of Markov chain (i.e. order can be higher than two).

Let $m_{i_0 i_1 i_2 \ldots i_n}$ be the expected number of observations passing through the sequence of states $i_0, i_1, i_2, \ldots, i_n$ where n is the maximum order to be tested, r is the number of states and

$i_x \in \{1, 2, \ldots, r\}$, $x=0, 1, 2, \ldots, n$.

The saturated model is

$$Log\ m_{i_0 i_1 i_2 \ldots i_n} = U + \sum_{j=0}^{n} U_{i_j} + \sum_{j=0}^{n-1} \sum_{k=j+1}^{n} U_{i_j i_k} + \sum_{j=0}^{n-2} \sum_{k=j+1}^{n-1} \sum_{l=k+1}^{n} U_{i_j i_k i_l} + \ldots + U_{i_0 i_1 i_2 \ldots i_n}$$

$U_{i_j i_{j+1} \ldots i_{j+k}}$ is the parameter representing the joint effect or interaction among states $i_0, i_1, i_2, \ldots, i_n$. For example, $U_{i_j i_k}$ represents the interaction between states i_j and i_k; $U_{i_j i_k i_l}$ stands for the interaction among states i_j, i_k, and i_l.

To illustrate this model, let us suppose that there are 2 states and the maximum order to be tested is 2 (r=2 and n=2), then

$$Log\ m_{111} = U + 3U_1 + 3U_{11} + U_{111}$$

$$Log\ m_{112} = U + 2U_1 + U_2 + U_{11} + 2U_{12} + U_{112}$$

$$Log\ m_{121} = U + 2U_1 + U_2 + U_{12} + U_{11} + U_{21} + U_{121} \qquad etc.$$

A stepwise approach to testing can eliminate interactions in decreasing order. The model that fits and has the lowest level of interactions is the best model and indicates the order of Markov model. The order of the Markov model equals the level of interactions of this model minus one. For example, if the best model is a four-way interaction model, then the order of the Markov model is three.

A sequence of at least n+1 moves (i.e., an n+1-tuple) is required to test an order of n. It is sufficient to use n+1 moves to test an order of n if the Markov chain is stationary. Therefore, if the maximum order to be tested is three, quadruple data (a sequence of four moves: move 1, move 2, move 3, move 4) should be used to test the log-linear model with three-way interactions. If this model fits, then triple data (a sequence of three moves: move 1, move 2, move 3) can be used to test the model with two-way interactions. It should be noted that the method stated here is different from Bishop et al's (1975, p. 269) approach. In Bishop et al's method, if the three-way model fits, then the quadruple data not triple data is still used to test two-way interactions. The advantage of my method is that the model to be tested is simpler, because only 2 two-way interactions (move 1 by move 2, move 2 by move 3) are needed for triple data, while 3 two-way interactions (move 1 by move 2, move 2 by move 3, move 3 by move 4) are needed for quadruple data. In short, this method states that the number of moves in the data sequence can decrease along with the decrease of the order to be tested.

In this study, the Markov model is assumed to be stationary. The software Statistical Package for the Social Sciences (SPSSx) on a VAX computer was used for this analysis. Considering the sample size and the number of states in the transition probability matrix, the maximum order tested was three. So, a log-linear model with the appropriate three-way interactions (i.e., move 1 by move 2 by move 3, move 2 by move 3 by move 4) was tested on quadruple data first. Because this model fit, the model with the appropriate two-way interactions (i.e., move 1 by move 2, move 2 by move 3) was then tested on triple data. Notice the word "appropriate" here. For triple data, there are three possible two-way interactions: move 1 by move 2, move 2 by move 3, and move 1 by move 3. However, only the first two interactions are appropriate for a first-order Markov model since the last interaction implies the influence of move 1 on move 3 which has a second-order nature and therefore inappropriate for a first-order Markov model.

Pearson chi-square statistic (X^2) was used as the criteria of goodness of fit instead of likelihood ratic chi-square statistic (G^2), because X^2 is preferable to G^2 for a small-sample matrix (Fienberg, 1980, p.173). Since there are a substantial number of structural zeros in the transition probability matrices (A structural zero at row i and column j in a transition probability matrix means it is impossible to have a transition from state i to state j), the degrees of freedom of the log-linear models were adjusted by subtracting the number of structural zeros (Knoke, 1980, p.65). For example, if the unadjusted degree of freedom from the SPSSx output is 100, and there are 20 structural zeros, then the adjusted degree of freedom will be 80, which will be used as the criteria for determining the appropriate P-value.

After determining the order of the Markov model, the transition probability matrix with that order was developed by maximum likelihood estimation in the following way. (Estimation for the second-order Markov model is given below since the Markov model in this study turned out to be second-order in nature.)

Suppose the Markov model has I possible states (in this study, I=8), and there are T sets of successive transitions from one state to another. Let $Y_{ijk}(t)$ be the number of observations in state i at time t-2, state j at time t-1, and state k at time t. Let $P_{ijk}(t)$ be the conditional probability of being in state k at time t, given state j at t-1, and i at t-2. For the stationary transition probability matrix, $P_{ijk}(t)=P_{ijk}$, for t=2, ... T. The maximum likelihood estimation of P_{ijk} is (Bishop et al, 1975, p. 268)

$$\hat{p}_{ijk} = \frac{Y_{ijk}}{\sum_{K=1}^{I} Y_{ijk}}$$

Where

$$Y_{ijk} = \sum_{t=2}^{T} Y_{ijk}(t)$$

Once the order, say 3, was determined, then additional parameters were added representing other factors (f), that is, different user groups

$$Log\ m_{i_0 i_1 i_2 i_3 f} = U + U_f + \sum_{j=0}^{3} U_{i_j} + \sum_{j=0}^{3} U_{i_j f} + \sum_{j=0}^{2} \sum_{k=j+1}^{3} U_{i_j i_k}$$

$$+ \sum_{j=0}^{2} \sum_{k=j+1}^{3} U_{i_j i_k f} + \sum_{j=0}^{1} \sum_{k=j+1}^{2} \sum_{l=k+1}^{3} U_{i_j i_k i_l} + \dots + U_{i_0 i_1 i_2 i_3 f}$$

The log-linear model tests were used to determine if there was a significant difference between different user groups. For the same reason stated above, X^2 instead of G^2 was used and the degrees of freedom were adjusted based on the number of structural zeros.

It should be noted that the four factors studied (gender, search task, search experience, and academic background) were tested one at a time (there is only one "f" in the above model instead of $f_1f_2f_3f_4$). That is, the above model assumes no interaction between or among different factors (e.g. interaction between gender and search experience). Sample size was not sufficient to explore these interactions. The Markov model in this study turned out to be second-order. The log-linear model would have seven dimensions if the four factors were tested together (added as four dimensions on top of the second order Markov model).

4.5.3 Analysis of Variance

Although different user groups' search patterns were compared by the Markov model, as discussed above, this part of the study aims at a direct comparison of analytical searching vs. browsing, while the Markov model comparison covers all search states including printing, getting help etc.. Information about users' search options resulting from this comparison will certainly benefit design decision. It can potentially suggest a future implementation that is suitable for different user groups and can support a range of information-seeking activities ranging from fact retrieval to general browsing. Such information can also be used in user training.

Three factors (gender, search experience, and search task) were studied for their influence on the search options. It was hypothesized that:

(1) There is no significant difference between males and females in terms of search option.

(2) Users with high online search experience tend to use more analytical searching than users with low search experience. This hypothesis was based on the assumption that the more search experience (presumably in Boolean logic based system) the user has, the more likely he/she will use analytical searching through habit.

(3) A general search task will lead to more browsing while a specific search task will lead to more analytical searching. The assumption is that browsing is inefficient for direct search task or fact retrieval (Marchionini & Shneiderman, 1988).

Based on the features of the experimental system (Hyperties version of Hypertext on Hypertext) used in this study, browsing was operationally defined as being in one of the following search states: searching through Table of Contents, searching through the Index, non-linear browsing, and linear browsing; while analytical searching was operationally defined as string searching (authors or keywords).

Since a search goes through a sequence of states, it is possible that a search consists of both analytical searching and browsing states. Therefore, the property of a search (analytical searching vs. browsing) was measured by the AB rate, which is defined as:

$$AB = \frac{\text{number of different nodes reached by analytical searching}}{\text{number of different nodes reached either by analytical searching or browsing}}$$

Notice that the AB rate is calculated based on the number of <u>different</u> nodes. Therefore, the nodes visited more than once in a search (resulting from the move of going back to a previously visited node) were counted only once. Help screens, search history, table of contents, index, and string searching nodes were not counted since they are the function nodes through which analytical searching or browsing is conducted.

A three-way analysis of variance (ANOVA) test was performed to test the hypotheses stated above. The dependent variable of the test was the AB rate and the independent variables were gender, search experience, and search task.

Because AB rate is proportion data, it is possible that one of the assumptions of the analysis of variance test, namely homoscedasticity, will be violated. (Judd & McClelland, 1989). The Bartlett-Box homogeneity test was therefore conducted before the analysis of variance tests. If there had been a violation of the homoscedasticity assumption, remedies such as arcsine transformation would have been applied.

Chapter 5. Results

5.1 The Markov Model

The log-linear model with the appropriate three-way interactions fits (X^2=1346.59, df=1310, p=0.24) and the log-linear model with the appropriate two-way interactions does not fit (X^2=1176.68, df=227, p<0.01). Therefore, the Markov model is of second order. This means that the past two moves influence the current move. Earlier moves have no influence on the current move. It should be pointed out that this does not mean that the user totally "forgets" what has been done two moves ago. Rather, it means that the current move can be predicted by the past two moves. The whole search process can be viewed as a sequence of moves with each move being influenced by the preceding two moves. For example, move 4 is influenced by move 2 and 3 while move 3 is influenced by move 1 and 2. When the user is at move 4, she/he does not "forget" what happened at move 1. Actually, move 1 contributed to move 3, which, in turn, contributed to move 4. Since the move 1 information is carried by move 3, we only need information about moves 2 and 3 to predict move 4. It is a chain of influence and moves, hence the term Markov chain.

Since the second-order Markov model has a three-dimensional transition probability matrix which can be written as eight two-dimensional transition probability matrices, it is rather lengthy and therefore presented in Appendix 3. The zero- and the first-order transition probability matrices are in Table 1 and Table 2 (expressed as percentages).

An "SZ" in the matrix stands for a structural zero. A structural zero at row i and column j means that it is impossible to have a transition from state i to state j.

Table 1 State (Zero-order Transition) Probability Matrix (%)

state	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
p bability	15.1	23.2	10.2	12.3	19.3	9.9	8.1	1.9

Table 2 First-order Transition Probability Matrix (%)

move 1 \ move 2	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
querying	12.2	80.9	0.9	3.9	SZ	SZ	SZ	2.2
displaying title	1.6	12.4	0.7	0.7	[illegible]	SZ	SZ	0.4
printing	22.8	23.9	SZ	23.0	0.7	8.9	18.9	1.8
node revisiting	25.0	10.0	5.1	17.3	SZ	19.4	17.1	6.0
displaying article	21.5	19.2	25.2	9.4	SZ	12.7	10.2	1.7
non-linear browsing	10.7	2.6	16.4	26.9	SZ	34.8	6.8	1.9
linear browsing	8.3	5.1	35.6	21.7	SZ	9.1	19.7	0.6
getting help	35.0	3.8	1.3	57.5	SZ	2.5	0.00	0.00

The zero-order Markov model shows the percent frequency of state occurrence rather than a transition occurrence. The state that occurred least frequently (1.9%) is GETTING HELP. Two actions were included in this state: using help screens and displaying path histories. Examination of the original transition data revealed that the help screen was called up only 14 times out of a total 4310 transitions, (0.32%). This low number is not surprising since most subjects in this experiment felt that the system was easy to use. Path History can be used as a reminder of the nodes already visited when the user feels disoriented. The low usage of the Path History in this experiment does not means users did not feel lost. In fact, quite a few subjects reported that they were lost, but they did not find the Path History to be particularly helpful since it is a "linear" list of the nodes visited rather than a map of the network. Some subjects said it would be a lot more helpful if the path and the current position were shown in a map form in relation to the global structure of the database. This result conforms with Edward and Hardman's conclusion from their study that "The most appropriate types of navigation devices would be those that are spatially-based, i.e. present the information structure in a 2 or 3-dimensional form, rather than those which simply keep account of the names of the screens that the user has viewed, ..." (Edward & Hardman, 1989, p. 123).

The most frequently occurring state is DISPLAYING TITLE (23.2%), followed by "DISPLAYING ARTICLE" (19.3%). Examining the first-order transition probability matrix, it is found that "DISPLAYING TITLE" state was followed predominantly by the state "DISPLAYING ARTICLE" (84.2%). This means after reading the title and the brief description (a couple of lines of description of the article will appear when the title is displayed), most users want to read the article itself. This could be attributed to the fact that a few lines of description of the content is not informative enough for users to make the relevance judgement, so they still have to see the article. For users (majority) who want to see the article instead of just the description of the article, it would be more efficient to display the article immediately (one move) instead of displaying the title first and then the article (two moves). So the system should provide the option of displaying articles directly (in the current system, an article cannot be displayed directly without displaying the description first).

The zero-order model shows that QUERYING state (including searching through Table of Contents, Keyword, and the Index) occurred 15.1% times, while NON-LINEAR BROWSING only 9.9% times. So, contrary to expectation, the non-linear browsing was not the predominant search strategy. It should be noted that there are more general task searches (thirty) than specific task searches (twenty-five) and the comparison of the state transitions of the two types of tasks shows that the general task group used more NON-LINEAR BROWSING than the specific task group did (see section 5.2.2). It can thus be reasoned that if the same number of searches were performed for each type of tasks, the proportion of NON-LINEAR BROWSING would be even lower.

The analysis of the second-order model revealed that the most frequently occurring triple in the state transitions is QUERYING -- DISPLAYING TITLE -- DISPLAYING ARTICLE (459 out of total 4310 transitions). The second most frequently occurring triple is: DISPLAYING TITLE -- DISPLAYING ARTICLE -- PRINTING (210 out of total 4310 transitions). Linking these two most frequently occurring triples results in the sequence of states: QUERYING -- DISPLAYING TITLE -- DISPLAYING ARTICLE -- PRINTING. This is a typical search pattern in traditional systems. It is clear that the predominant search strategy adopted is not very different from what is used in traditional Boolean systems. Hypertext search features were used as an auxiliary tool. For example, the triple NON-LINEAR BROWSING -- NON-LINEAR BROWSING -- NON-LINEAR BROWSING occurred in only 92 out of 4310 transitions.

This result has an important implication for future design:

keyword and Boolean search function should be fully implemented even in a hypertext system, since it will be a major feature used. Inadequate implementation of this function will cause user dissatisfaction. In fact, a lot of subjects complained about the slow speed of keyword search in this experimental system (it appears the system does not have a keyword inverted file). The slow speed of keyword search deterred, to a great extent, the use of the strategy: keyword searching -- displaying title -- displaying article -- printing. Otherwise, this pattern would likely be even more predominant.

Since most subjects in this study are library and information science students, it is possible that their searching patterns are coloured by their experience with other retrieval systems. However, a study of a hypertext browsing system in which the subjects were not library and information science students also revealed that keyword search was a common and effective strategy (Egan et al., 1989). Some subjects in that study had no experience in information retrieval systems. Another study, conducted by Marchionini et al. (1990), also indicates that non-linear browsing was used much less than expected. Even though string searching is particularly limited in that system, it still was the dominant strategy chosen by the experts. Walker et al. (1990) reported a study on an online full text database. Their results indicate very little table of contents usage and heavy use of search commands. They thus reasoned: "The heavy reliance of users on searching to find information should suggest to documentation writers that well-chosen keywords are even more important for an online manual than for printed books." Halasz (1987) discusses several enhancements which he believes are critical for the success of the next-generation hypermedia systems. Among the most important of these suggestions is that link-based navigation methods be supplemented with content-based mechanisms. That is, instead of having to navigate through the network, users should be able to go directly to notecards of interest by specifying keywords.

One of the possible reasons for the heavy usage of keyword searching is users' habit in using printed books. Locating relevant information through the back of the book index is similar to keyword searching in a database -- accessing information through specific word(s). This explains why users who have little searching experience also tend to use keyword searching. Their experience in using book index leads them to conduct keyword searching although they have little experience in Boolean searching.

In hypertext systems, a single paper is cut into several small "chunks" and stored in different nodes. The assumption is that instead of reading the same paper linearly, users would like to jump from one paper to another. However, the zero-order Markov model shows 8.1% occurrence of the state LINEAR BROWSING (moving forward or backward in the same paper) which is only a little lower than the occurrence of the state NON-LINEAR BROWSING (9.9%). Since NON-LINEAR BROWSING is jumping to another node by clicking on a highlighted term in the text, it is possible that the two nodes belong to the same paper. So, 9.9% is the maximum probability of moving to a <u>different</u> paper by NON-LINEAR BROWSING.

It is further noticed that there is a 19.7% probability of moving from the state LINEAR BROWSING to LINEAR BROWSING. There seems to be a tendency to continue reading the same paper. Notice the relative high conditional probability of moving from PRINTING to LINEAR BROWSING (18.9%). Once one part of a paper is considered to be relevant and printed, then other parts of the same paper are considered to be possibly relevant and worth examining.

5.2 Comparison of Different User Groups Using Markov Models

A transition probability matrix describes in matrix form the probability of going from one state to another state. The pattern of movement through the states can be used as a map of the user behaviour. The search patterns of different user groups can thus be compared by examining the corresponding transition probability matrices. The following user groups were compared in this study: gender, search experience, search task, and subject's academic background (academic discipline chosen by the subject, that is, science or engineering vs. social science or humanities).

The log-linear model test was used to determine if there is a significant difference between groups. If there is a difference, then the corresponding transition probability matrices were examined to determine the pattern of difference. The log-linear model tests were conducted on the second-order transition probability matrices since the test of the order turned out to be two. However, first-order models were compared to see how the two groups differ, since the second-order models have eight two-dimensional matrices which are too long to be compared conveniently. The results of these comparisons are reported below.

5.2.1 Gender

Borgman found that males and females perform differently in a Boolean logic based system (see Chapter 3.4). However, there is no report as to how they differ.

In this study, comparison of male vs. female behaviour also turned out to be significant (X^2=365.50, df=181, p<0.01). The examination of the zero-order transition probability matrices (Table 3) revealed no major difference in overall usage of each state. The differences come from the transition of states or the patterns of moves. The comparison of the first-order transition probability matrices (Table 4) shows that males and females differ mainly in the following aspects.

Table 3 State (Zero-order Transition) Probabilities for Different Gender Groups (%)

There were twenty male subjects and thirty-five female subjects.

state / gender	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
male	13.8	24.4	9.8	11.2	20.7	10.3	7.3	2.4
female	15.6	22.7	10.3	12.8	18.7	9.8	8.5	1.6

Table 4 First-order Transition Probabilities for Different Gender Groups (%)

In each cell, the figure on the first line is for male group and the second line for female group.

There were twenty male subjects and thirty-five female subjects.

move 1 \ move 2	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
querying	12.6 12.0	76.6 82.5	0.6 1.1	5.1 3.4	SZ	SZ	SZ	5.1 1.1
displaying title	2.9 1.0	9.1 13.9	0.00 1.0	0.3 0.9	87.4 82.8	SZ	SZ	.3 0.4
printing	16.1 25.4	31.5 21.0	SZ	25.0 22.2	0.00 1.0	9.7 8.6	14.5 20.6	3.2 1.3
node revisiting	25.4 24.9	15.5 8.0	4.9 5.1	10.6 19.8	SZ	16.2 20.6	21.1 15.7	6.3 5.9
displaying article	17.6 23.3	26.3 16.0	21.4 27.0	8.4 9.8	SZ	13.4 12.5	10.3 10.2	2.7 1.2
non-linear browsing	10.0 11.1	1.5 3.0	17.7 15.8	23.8 28.2	SZ	37.7 33.6	6.9 6.7	2.3 1.7
linear browsing	7.5 8.5	10.8 3.1	40.9 33.7	20.4 22.1	SZ	9.7 8.9	10.8 22.9	0.00 0.8
getting help	35.5 34.7	3.2 4.1	0.00 2.0	54.8 59.2	SZ	6.5 0.00	0.00 0.00	0.00 0.00

There is a higher probability for females to move from NODE REVISITING to NODE REVISITING (females 19.8% vs. males 10.6%). Females are more likely, than males, to linear browse continuously, (probability of LINEAR BROWSING TO LINEAR BROWSING is 22.9% for females and 10.8% for males). These kind of moves (keep tracing back to previous nodes, or moving backward or forward in a paper) are more likely to cause a feeling of being lost than keyword searching, because of the former's unstructured nature. Using more these kind of moves would indicate that females are less concerned with disorientation. It is difficult to explain why females are more likely to keep tracing back to previous nodes. It might mean that they remember the previous nodes better and therefore more likely to go back to revisit them. On the other hand, it could also mean that they tend to forget the information contained in the previous node and therefore have to go back to read it again.

There is a higher probability for males than females to move from DISPLAYING ARTICLE to DISPLAYING TITLE (male 26.3%, female 16%) and from DISPLAYING TITLE to DISPLAYING ARTICLE (male 87.4%, female 82.8%). Males also have a higher probability of going into these two states to start with (see zero-order model). After printing an article, males have a higher probability of displaying another article (male 31.5%, female 21%).

To summarize, males are more likely than females to adopt the search strategy: conduct a keyword search, and then examine the search result one after another by repeating the cycle, displaying title -- displaying article -- or printing if the article is relevant. Notice that this is a typical search strategy applied in Boolean logic based systems. So males tend to use more traditional strategy than females, and females tend to do more continuous linear browsing and node revisiting, but not more continuous NON-LINEAR BROWSING.

5.2.2 Task

User studies of information retrieval systems often control the factor of task. Borgman (1986) compared the user performance between simple task and complex task. Marchionini (1989) studied the difference between open task and closed task.

In this study, the tasks compared are general task and specific task (see Chapter 4.2 for the description of the tasks). The log-linear model test showed that the search patterns of the users with these two different types of tasks are significantly different (X^2=957.53, df=181, $p<0.01$). See Table 5 and Table 6 for the comparison of zero- and first-order transition probability matrices.

The differences between the two task groups are very clear even in the zero-order matrix. Browsing and node revisiting states occurred more often in the general task group. The discrepancies are 14.4% vs. 8.7% for NODE REVISITING; 11.8% vs. 6.6% for NON-LINEAR BROWSING; and 9.1% vs. 6.5% for LINEAR BROWSING. These states use hypertext system rather than traditional system features. In the specific task group, states DISPLAYING TITLE and DISPLAYING ARTICLE occurred more (30.1% vs. 19.3% for DISPLAYING TITLE; 24.5% vs. 16.4% for DISPLAYING ARTICLE). The QUERYING state includes searching through keyword, Table of Content, and the Index, so it is not clear from the transition probability matrices what kind of search is used in the specific task. An examination of the original data revealed that the specific task group conducted more keyword searches than the general task group. This is confirmed in the next part of this thesis, analysis of variance test.

Table 5 State (Zero-order Transition) Probabilities for Different Task Groups (%)

Thirty subjects performed general task and twenty-five subjects performed specific task.

state / task	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
general task	16.8	19.3	10.4	14.4	16.4	11.8	9.1	1.9
specific task	11.9	30.1	9.8	8.7	24.5	6.6	6.5	1.8

Table 6 First-order Transition Probabilities for Different Task Groups (%)

In each cell, the figure on the first line is for general task group and the second line for specific task group.

Thirty subjects performed general task and twenty-five subjects performed specific task.

move 1 \ move 2	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
querying	9.9 17.7	83.6 74.2	1.1 0.5	3.5 4.8	SZ	SZ	SZ	1.9 2.7
displaying title	1.5 1.7	9.8 15.3	0.8 0.6	0.9 0.4	86.2 81.9	SZ	SZ	0.8 0.00
printing	29.0 11.1	8.4 52.9	SZ	29.7 10.5	0.00 2.0	9.8 7.2	21.0 15.0	2.1 1.3
node revisiting	25.6 23.5	6.6 19.9	4.3 7.4	18.7 13.2	SZ	20.5 16.2	19.0 11.8	5.3 8.1
displaying article	27.8 14.1	6.7 34.0	27.1 23.0	10.7 7.9	SZ	16.0 8.9	10.4 9.9	1.3 2.1
non-linear browsing	11.1 9.7	1.2 6.8	16.9 14.6	25.2 32.0	SZ	37.2 27.2	6.5 7.8	1.8 1.9
linear browsing	8.0 8.8	2.0 12.7	34.9 37.3	24.9 13.7	SZ	10.0 6.9	19.7 19.6	0.4 1.0
getting help	38.5 28.6	1.9 7.1	0.00 3.6	59.6 53.6	SZ	0.00 7.1	0.00 0.00	0.00 0.00

The comparison of the first-order Markov model further revealed the different patterns of the two task groups. The general task group not only has a higher probability of non-linear browsing and node revisiting to start with, but also has a higher probability of doing it continuously. The probability of moving from NON-LINEAR BROWSING to NON-LINEAR BROWSING is 37.2% compared to 27.2% for a specific task; the probability of going from NODE REVISITING to NODE REVISITING is 18.7% compared to 13.2% for a specific task. In contrast, the specific task group has a higher probability of moving from DISPLAYING ARTICLE to DISPLAYING TITLE (34% vs. 6.7%), and from PRINTING to DISPLAYING TITLE (52.9% VS. 8.4%). This shows that the specific task group repeated the circle DISPLAYING TITLE -- DISPLAYING ARTICLE -- PRINTING once a keyword searching resulted a set of hits. This is a typical search strategy of traditional Boolean systems.

The specific task group also has a higher probability of moving from NON-LINEAR BROWSING to NODE REVISITING (32% vs. 25.2%). This means that the specific task group will not wander too far away. After moving one step away from the current node, they are most probably going to come back immediately. The words "most probably" are used here, because for the specific task group, the state that has the highest probability following NON-LINEAR BROWSING is NODE REVISITING.

The difference of the two search task groups can be summarized as follows. For specific tasks, users tend to adopt the structured search pattern: keyword searching -- displaying title -- displaying article. For general task, users tend to wander from node to node browsing without a structured path pattern. A general task caused users to use more hypertext system browsing features while specific tasks caused users to use more traditional system features. It appears that hypertext system features are more suitable for general browsing while the traditional Boolean logic system features are more appropriate for specific queries.

When Marchionini et al. found the unexpected result that searchers used much more query searching and indexing than non-linear browsing, they stated: "Whether these effects are due to the nature of the search tasks, the system itself, or subjects' inexperience with the hypertext medium is an important issue for future research since much of the interest in hypertext relates to the ability to browse freely and rapidly for information" (Marchionini et al., 1990, p.133). The result of this study answered one of the questions they raised, namely the nature of the search tasks. As to the system factor, it does not seem to be the cause. This study used the Hyperties system. Marchionini et al.'s study used the HyperCard system. Egan et al.'s study used the SuperBook system. All three studies found that keyword searching was the main strategy used. The question of whether the subjects' inexperience with the hypertext medium is a factor in search patterns is indirectly explored in this study through the comparison of search experience reported below.

5.2.3 Search Experience

Most information retrieval experiments control the factor of search experience prior to the study. In this study, very few subjects had experience in hypertext systems, but their search experience with other information retrieval systems varied. Subjects were asked to rate themselves for their amount of search experience (see Chapter 4.4). Two levels of search experience (high experience vs. low experience) were compared.

The log-linear model test indicates that there is a significant search pattern difference between users with different experience (X^2=305.20, df=181, $p<0.01$). See Table 7 and Table 8 for the comparison of transition probabilities of the two experience groups.

Table 7 State (Zero-order Transition) Probabilities for Different Experience Groups (%)

Thirty-three subjects rated themselves as having low experience and twenty-two high experience.

state / experience	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
low experience	14.3	23.4	11.1	11.4	19.6	9.8	8.8	1.6
high experience	16.2	22.9	8.9	13.6	18.9	10.2	7.2	2.2

Table 8 First-order Transition Probabilities for Different Experience Groups (%)

In each cell, the figure on the first line is for low experience group and the second line for high experience group.

Thirty-three subjects rated themselves as having low experience and twenty-two high experience.

move 1 \ move 2	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
querying	11.5 13.0	81.6 80.0	0.8 1.1	4.1 3.5	SZ	SZ	SZ	1.9 2.5
displaying title	1.3 2.0	12.7 11.9	0.8 0.5	0.7 0.7	84.1 84.3	SZ	SZ	0.3 0.5
printing	17.3 32.7	30.0 12.8	SZ	22.3 24.4	1.1 0.00	8.1 10.3	19.1 18.6	2.1 1.3
node revisiting	25.7 24.3	8.6 11.7	6.5 3.3	14.4 20.9	SZ	22.6 15.5	17.1 17.2	5.1 7.1
displaying article	19.6 24.3	19.0 19.5	25.5 24.9	8.0 11.4	SZ	13.8 11.1	12.4 6.9	1.6 1.8
non-linear browsing	12.4 8.4	3.2 1.7	18.1 14.0	28.1 25.1	SZ	29.3 42.5	7.6 5.6	1.2 2.8
linear browsing	9.3 6.3	4.0 7.1	39.1 29.4	19.1 26.2	SZ	8.4 10.3	19.1 20.6	0.9 0.00
getting help	41.5 28.2	4.9 2.6	0.00 2.6	51.2 64.1	SZ	2.4 2.6	0.00 0.00	0.00 0.00

Comparing the two matrices, the major difference between the two groups is that the high experience group seems to have a higher probability of non-linear browsing and back tracing nodes continuously. Probability of moving from state NON-LINEAR BROWSING to NON-LINEAR BROWSING is 42.5% for high experience users vs. 29.3% for low experience users; and probability of moving from state NODE REVISITING to NODE REVISITING is 20.9% for the former and 14.4% for the later. In contrast, low experienced users applied more LINEAR BROWSING than high experience users (Zero-order transition probability matrices show a 8.8% chance of LINEAR BROWSING for low experience users and a 7.2% chance for high experience users). When low experience users were in the NON-LINEAR BROWSING state, they have a higher probability of changing to QUERYING (probability of moving from NON-LINEAR BROWSING to QUERYING is 12.4% for low experience group compare to 8.4% fro high experience group).

To summarize, high experience users tend to use hypertext features moving from node to node non-linearly and later tracing back continuously while low experience users tend to be more linear or structured in their path pattern. They appear to feel more in control with a linear searching path.

One possible explanation for this phenomenon is that experienced users feel more confident and relaxed in computerized retrieval systems, so that they can wander away from the current node freely and later trace back continuously without being too concerned with becoming lost in the hypertext space.

Although the experience level being compared here refers to subjects' experience with non-hypertext systems, it could be hypothesized that subjects' experience with hypertext medium will influence their search strategy too. Perhaps the more experience subjects have with the hypertext medium, the more they tend to browse non-linearly. This is an interesting and important question worthy of further study.

5.2.4 Subject's Academic Background

The subject's academic background refers to the academic discipline studied by the subject, (i.e., science, engineering, social science, or humanities). In order to have enough data points to conduct the appropriate statistical test, subjects with a science or engineering background were combined to form a technical group while subjects with social science or humanities background were combined to form a non-technical group.

The log-linear model test indicates that there is a statistically significant difference between the two groups (X^2=242.40, df=181, p=0.002). The zero- and first-order transition probability matrices for the two groups are presented in Table 9 and Table 10.

Non-technical group users tend to trace back further the nodes already visited (probability of moving from NODE REVISITING to NODE REVISITING is 17.8% for the non-technical group compare to 11.6% for the technical group). The technical group users have a higher probability of following hypertext links continuously (44.7% vs. 33.6% moving from NON-LINEAR BROWSING to NON-LINEAR BROWSING). Interesting enough, they are less likely to find relevant articles through browsing. The probability of moving from NON-LINEAR BROWSING to PRINTING is 17.3% for non-technical group but only 8.5% for technical group. The probability of moving from LINEAR BROWSING to PRINTING is 36.5% for non-technical group and 27.8% for technical group. The assumption here is that printing means finding relevant articles.

Table 9 State (Zero-order Transition) Probabilities for Different Background Groups (%)

Forty-eight subjects had non-technical background and seven subjects had technical background.

state / academic background	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
non-technical	15.1	23.2	10.3	12.5	19.2	9.7	8.1	1.9
technical	14.5	23.4	9.2	10.7	20.4	11.7	9.0	1.0

Table 10 First-order Transition Probabilities for Different Background Groups (%)

In each cell, the figure on the first line is for non-technical group and the second line for technical group.

Forty-eight subjects had non-technical background and seven subjects had technical background.

move 1 \ move 2	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
querying	11.8 15.5	80.7 82.8	1.0 0.00	4.2 0.00	SZ	SZ	SZ	2.2 1.7
displaying title	1.5 2.1	13.1 5.3	0.8 0.00	0.6 2.1	83.6 90.4	SZ	SZ	0.4 0.00
printing	22.9 21.6	23.9 24.3	SZ	22.6 27.0	0.7 0.00	9.2 5.4	18.7 21.6	2.0 0.00
node revisiting	25.0 25.6	10.0 9.3	4.9 7.0	17.8 11.6	SZ	20.1 11.6	16.2 27.9	5.9 7.0
displaying article	21.9 18.3	18.5 25.6	25.2 25.6	9.6 7.3	SZ	12.3 17.1	10.7 6.1	1.9 0.00
non-linear browsing	11.0 8.5	2.6 2.1	17.3 8.5	27.6 21.3	SZ	33.6 44.7	6.0 12.8	1.8 2.1
linear browsing	8.6 5.6	4.8 8.3	36.5 27.8	21.3 25.0	SZ	8.3 16.7	20.0 16.7	0.6 0.00
getting help	34.2 50.0	2.6 25.0	1.3 0.00	59.2 25.0	SZ	2.6 0.00	0.00 0.00	0.00 0.00

The difference between these two groups is not as clear cut as that of other groups and it is difficult to summarize the difference pattern. The small sample size for the technical group (seven subjects) makes me reluctant to draw any firm conclusions and interpretations. Further research should seek to have a more balanced sample and explore this issue further.

5.3 Comparison of Search Options (analytical searching vs. browsing)

Table 11 provides the ANOVA result for the comparison of the two search options. Table 12 lists the mean AB rate values and the number of subjects in each category. As is indicated by the three-way analysis of variance test result, there is a statistically significant difference between the two task groups (p=0.001). As was hypothesized, the specific task group used more analytical search while the general task group browsed more. The mean of AB rate is 0.51 for the specific task group and 0.16 for the general task group.

Table 11 Result of the Three-way ANOVA

Factor	DF	F	P
Gender	1	0.95	0.334
Experience	1	0.47	0.495
Task	1	12.53	0.001
Gender by Experience	1	3.10	0.085
Gender by Task	1	1.06	0.309
Experience by Task	1	0.00	0.987
Gender by Experience by Task	1	1.26	0.268

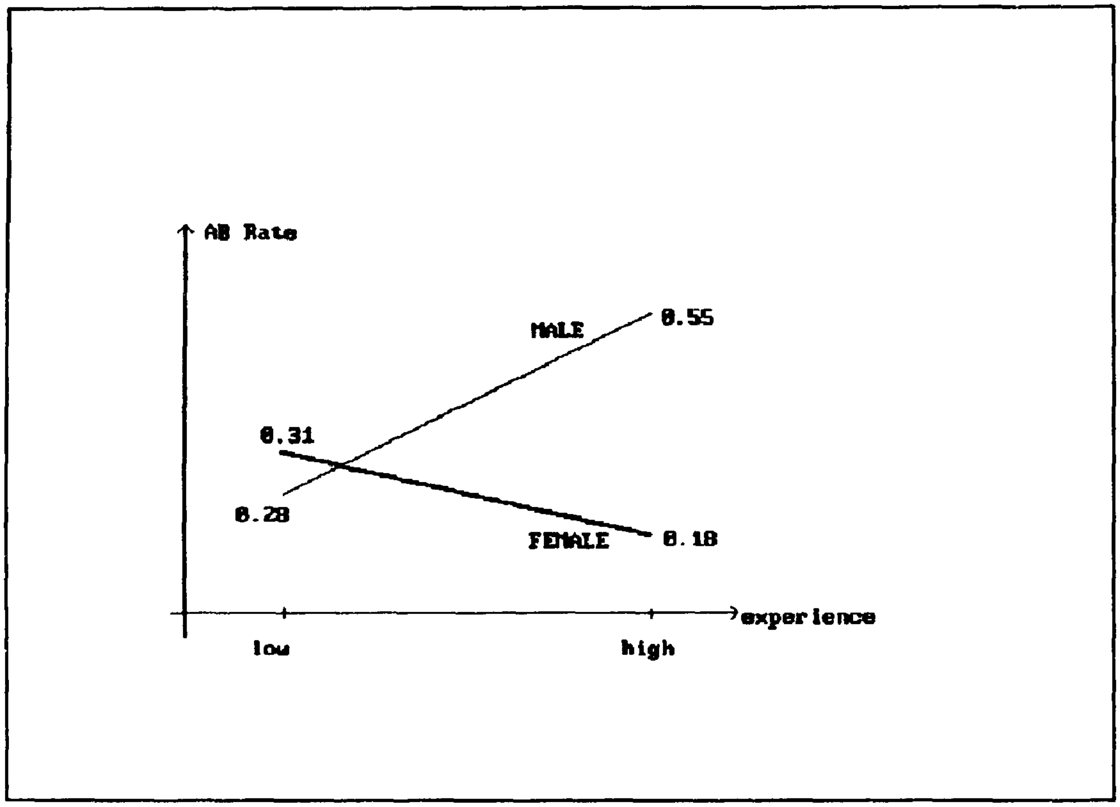
Table 12 Mean AB rates

Factor	Mean	Sample Size
male, low experience, general task	0.108	4
male, low experience, specific task	0.394	6
male, high experience, general task	0.217	3
male, high experience, specific task	0.692	7
female, low experience, general task	0.166	12
female, low experience, specific task	0.467	11
female, high experience, general task	0.169	11
female, high experience, specific task	0.286	1

No other factors and interactions of factors were revealed to be significant by the three-way ANOVA at the 0.05 significance level. However, attention was drawn to the fact that there might be a two-way interaction between factors of gender and search experience (p=0.085). Since one cell in the three-way design contained only one observation, this three-way ANOVA result should be taken very cautiously. Besides, the Bartlett-Box homogeneous test was not conducted for the same reason.

In order to further investigate the interaction between gender and search experience, a two-way analysis of variance test was carried out. Although a two-way ANOVA test is not very appropriate when three factors are under investigation, the absence of the three-way interaction makes this issue of less concern. In fact, this a practical solution to this dilemma, provided that the three- and two-way ANOVA results are interpreted carefully and appropriately.

FIGURE 7. Mean AB Rate Values



The Bartlett-Box test was conducted first and there was no evidence that the homogeneity assumption was violated (p=0.356). A two-way ANOVA test was then conducted. Figure 7 presents the mean AB rate values for each combination of factors. As was expected, following the three-way ANOVA result, there was an significant interaction between factors gender and experience (p=0.011). The crossing of the two gender lines at Figure 7 shows this interaction. A somewhat surprising result was that there was a significant gender difference (p=0.029). Males, in general, tended to use more analytical search than females did (mean of AB rate is 0.42 for males and 0.26 for females). There was no evidence that search experience is a main effect (p=0.366). In other words, search experience alone did not influence search option.

The significant interaction means that the gender effects are not the same at different levels of search experience. Among users with low experience, males were less likely to adopt analytical search than females did. On the other hand, for users with high experience, males tended to use more analytical search than females did. However, the small distance between the two gender lines at the low experience end suggests that the gender difference for low experience users may not be significant.

To further study the pattern of the gender difference, an analysis of simple main effects was carried out. As a matter of fact, this analysis should be carried out whenever the interaction is significant (Keppel & Saufley, 1980). The analysis of simple main effects tests the significance of one factor when another factor is controlled at a specific level. In this study, the significance of gender factor was tested at the low experience level and the high experience level separately.

The result of the analysis of simple main effect indicates, as was expected, that the gender difference is not significant for low experience users ($p=0.949$). But it is significant for high experience users ($p=0.002$) with males using more analytical searching.

Summarizing the results of all these tests, it can be concluded that:

(1) Search task influences users' search option. A general search task leads to browsing while a specific search task leads to an analytical search.

(2) Search experience alone does not have an effect on the search option.

(3) There may be a gender difference (although there is no evidence in the three-way ANOVA). If this difference exists, it presents only among high experience users, with males being more in favour of analytical searching than females.

(4) The only possible interaction is between the factors of gender and search experience. This had borderline significance (p is 0.085 for the three-way ANOVA and 0.011 for the two-way ANOVA). There was no other interaction between or among factors.

It is not difficult to understand the search option differences resulting from the difference in search tasks. The search topics for specific tasks were relatively narrow, so following through links in an unstructured fashion may not provide many answers. On the other hand, the general task required a broader coverage of the topic and non-linear browsing can serve this purpose better.

It is difficult to interpret the pattern of gender difference, in particular, why the gender difference did not exist among low experience users. Perhaps low experience users were not "mature" enough in searching that their real characters would show in their searches. For high experience users, why are males more likely to adapt analytical search than females? There is no simple answer to this question. Besides, the result is not clear cut. There is no evidence in the three-way ANOVA result, and considering the unequal sample size in each cell, the difference in the two-way ANOVA result could be attributed to the difference in the task factor. Therefore, any interpretation would be premature or misleading. This question remains to be investigated in future research with larger and more balanced sample size.

There are two directions in which this study can be pursued further. The first is to conduct an additional analysis of variance of user groups in which the obviously nonsignificant interactions are removed from the three-way model and a simpler model fitted. The simpler model would contain only the main factors GENDER, EXPERIENCE, TASK and the board line significant interaction GENDER BY EXPERIENCE, thus allowing investigation of the interaction in the presence of the factor TASK. The other direction for further study is to use a log-linear or logistic-linear model instead of the three-way ANOVA to make the search option comparison. Because the outcome of the search option is a number of occurrences (number of nodes reached either by analytical searching or browsing), a log-linear or logistic-linear model can be applied. For example, the log-linear model can be a four dimensional model with m_{ijkl} being the expected number of observations for the ith task, the jth gender group, the kth experience group, and the lth search option (analytical searching or browsing). It would be interesting to compare the ANOVA result with the log-linear or logistic-linear model result.

5.4 Path Length Distribution

In this study, a path is defined as a sequence of nodes visited during a search. The path length is defined as the number of nodes in the path. It was hypothesized that most paths would be of middle length while a minority of paths would be very long or very short. The frequency distribution of data collected confirmed this hypothesis. The possible discrete models for this distribution would be: Poisson distribution, binomial distribution, and negative binomial distribution.

Since the path length cannot be zero, either a shifted or truncated version of these models should be used. The parameter estimations for the truncated version of these models are very complicated (Johnson & Kotz, 1969, p.73, p.105, and p.136). There are no simple solutions to the equations. Therefore, the shifted models were used.

The path lengths of the fifty-five searches were determined and the resulting distribution is listed in Table 13. These original data are grouped and presented in Table 14 for the convenience of examination. Notice that the path length data for the two search tasks (general task and specific task) were combined here because the two groups of data are homogeneous in terms of path length. The general task group has a mean path length of 39.67 with a standard deviation of 19.49 while the specific task group has a mean of 39.36 with a standard deviation of 21.03. A T-test comparing the path lengths of the two task groups did not reveal a significant difference ($p=0.96$). The type of task does not seem to influence the path length. So the two groups of data were combined for model fitting and testing. The mean of the path length for the combined data is 39.53 with a standard deviation of 20.01. Fitting and testing results are as follows. (In the following, $\overline{X}$ is sample mean and s^2 is sample variance)

Table 13 Path Length Data

X -- path length

Y -- number of searches with path length X

X	5	7	13	15	16	17	20	22	23
Y	1	1	2	1	1	4	1	2	3

X	25	26	28	29	30	31	34	35	37	38
Y	2	1	1	1	1	3	1	2	1	1

X	39	43	48	49	51	53	55	57	58	59
Y	1	1	2	3	1	1	1	1	1	2

X	63	64	65	66	68	71	72	75	76
Y	1	2	1	1	2	1	1	1	1

Table 14 Grouped Path Length Data

X	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80
Y	2	9	11	9	6	7	7	4

5.4.1 Shifted Binomial Distribution

The probability function for the shifted binomial distribution is

$$P(X) = \binom{n}{X-1} p^{X-1} \, (1-p)^{n-(X-1)}$$

where 0<p<1, n is a positive integer, for X=1,2,...n, n+1
The moment estimates of p and n are given by

$$\hat{p} = 1 - \frac{s^2}{\overline{X} - 1}$$

$$\hat{n} = \frac{\overline{X} - 1}{\hat{p}}$$

The path length data in this study has the property of $s^2 > \overline{X}-1$ which resulted in a $\hat{p} < 0$, indicating that shifted binomial distribution is not an appropriate model.

5.4.2 Shifted Poisson Distribution

The probability function for shifted Poisson distribution is

$$P(X) = \frac{e^{-\lambda} \lambda^{X-1}}{(X - 1)!}$$

where $\lambda > 0$ for X=1, 2, ... The maximum likelihood estimate of λ is

$$\hat{\lambda} = \bar{X} - 1$$

With $\hat{\lambda} = 38.53$, both goodness of fit tests (chi-square test and K-S test) failed ($X^2 = 161.52$, larger than the critical value of 14.07 with df=7. K-S=0.361, larger than the critical value of 0.215 for n=38). So the shifted Poisson distribution is not an appropriate model.

5.4.3 Shifted Negative Binomial Distribution

The probability function for shifted negative binomial distribution is:

$$P(X) = \binom{V+X-2}{X-1} p^V (1-p)^{X-1}$$

where $0 < p < 1$, V is a positive integer, for X=1,2, ... The moment estimates of p and V are given by

$$\hat{p} = \frac{\bar{X}-1}{S^2}$$

$$\hat{V} = \frac{\hat{p}\,(\bar{X}-1)}{1-\hat{p}}$$

The result of parameter estimation is $\hat{p} = 0.096$ and $\hat{V} = 4.1$. $\hat{V}$ was rounded to integer 4 and $\hat{p}$ was then adjusted accordingly which turned out to be 0.094. The model passed both chi-square test and K-S test. (For chi-square test, X^2 = 10.49 which is less than the critical value of 14.07 with df=7. K-S value is 0.101 which is smaller than the critical value of 0.215 for n=38.) When the $\hat{p}$ value was not adjusted after $\hat{V}$ value was rounded to integer 4, the model also passed both chi-square and K-S test but with a higher X^2 and K-S value (X^2=11.27, K-S=0.12). It can be predicted that if the $\hat{V}$ value was 4.5 and the $\hat{p}$ was not adjusted based on the change of $\hat{V}$ value, the model might not pass the goodness of fit test. So, it is recommended that the $\hat{p}$ value be adjusted after $\hat{V}$ value is rounded to an integer, especially when the $\hat{V}$ value lies between two integers.

In conclusion, the shifted negative binomial distribution is the appropriate model to describe the path length distribution. The result that only the shifted negative binomial model fits among the three models tested is not surprising. The most obvious distinction among the three types of distribution is in the values of the ratio of variance to mean. This is less than 1 for the binomial, equal to 1 for the Poisson, and greater than 1 for the negative binomial distribution (Johnson & Kotz, 1969). For shifted models, it is the ratio of variance to mean minus one [$s^2/(\bar{X}-1)$] that distinguishes the three models. In this data set, ratio of variance to mean minus one is

$20.01^2/38.53$, far greater than 1, thus resulting a fit of the shifted negative binomial model.

5.5 Frequency Distribution of Node Visiting

In hypertext systems, users can go back to a previously visited node after visiting other nodes. Thus, a single node can be visited more than once during a single search. Suppose X represents the number of times a node is visited during a search, and Y represents the number of nodes which were visited X times. The frequency distribution of node visiting is defined as Y = f(X). It was conjectured that this distribution would be the Zipf type, which means that the distribution will be in a hyperbolic curve described by the model

$$Y = \frac{A}{X^B}$$

where A and B are the parameters.

For each of the fifty-five search logs collected in this study, the number of nodes were counted for each frequency of X (e.g. how many nodes were visited once, how many twice, etc.). The result was presented in Table 15.

Table 15 Data of Frequency of Node Visiting

number of visits	1	2	3	4	5	6	7	8
search 1	18	4	1					
search 2	15		1	2				
search 3	5	1		1				
search 4	26	3	1			1		
search 5	7	3						
search 6	19	3	2	1				
search 7	31	5	2					
search 8	20	1	1	1				
search 9	7	1	2					
search 10	9	3	2					
search 11	6	2	1					
search 12	5							
search 13	29	5	2			1		
search 14	10					1		
search 15	12	10	3	1		1		
search 16	10	3	2					
search 17	17	3	1	2				
search 18	7	1	1	1				
search 19	31	10	1					
search 20	12	4						
search 21	11	1						
search 22	10	3						
search 23	6	2						
search 24	16	5	5			2		
search 25	16	5			2		1	
search 26	22	2						
search 27	25	8						

search 28	16	6	1	1				
search 29	28	7	1		3			
search 30	17	3			1			
search 31	6	1						
search 32	2	1						
search 33	10	3			1			
search 34	5	2						
search 35	10	2	1					
search 36	9	3						
search 37	16	3	1		1			
search 38	17	4	2	1				
search 39	27	4	1	1				1
search 40	8	3	1	1				
search 41	7	3						
search 42	5	1		1				
search 43	22	2	1			1		
search 44	33		1					
search 45	35	3						
search 46	13	1	2					
search 47	6	2						
search 48	19		2					
search 49	20	3		1				
search 50	21	2	3					
search 51	29				1			
search 52	9	2						
search 53	26	2	1	1				
search 54	15							
search 55	11	3	1			1		
Total	**844**	**154**	**47**	**16**	**9**	**8**	**1**	**1**

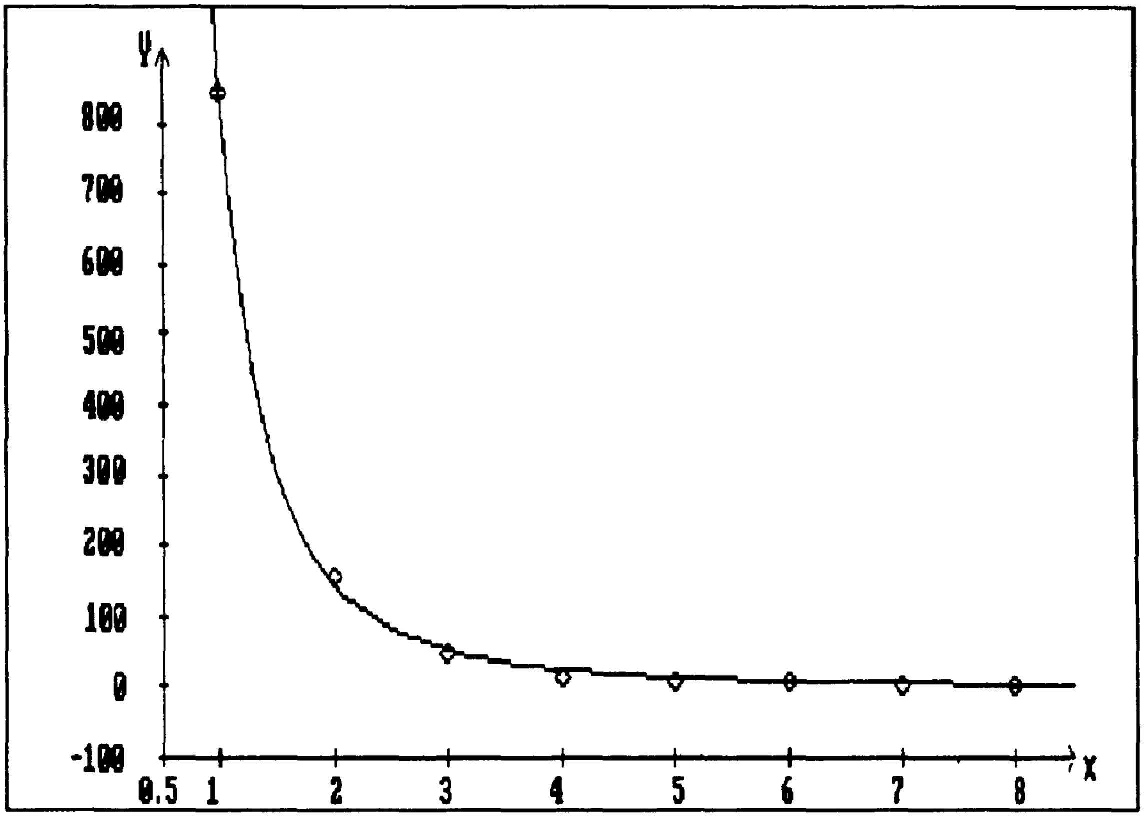
It should be noted that the following nodes were excluded from counting: table of content node, index node, help screen node, path history node, and the node containing the string search result. The reason for excluding these nodes is that they are the "function" nodes of the system through which searches were conducted. Because of their special function, they were visited a lot more than regular text or graphic nodes. Apparently, they will not follow the same frequency distribution and should be excluded so that the frequency distribution data represent a homogeneous population.

Frequency data of the fifty-five searches were cumulated at the bottom line of Table 15. It was these cumulated data that were used to fit the Zipf distribution. They are re-listed as the first two lines of Table 16 for the convenience of discussion. Data in Table 15 presents a very clear pattern: as X value goes up, Y value goes down dramatically. When these data points were plotted in Figure 8, a very clear hyperbolic curve formed.

Table 16 Observed vs. Expected Data for the Zipf Distribution

X value	1	2	3	4	5	6	7	8
observed Y value	844	154	47	16	9	8	1	1
expected Y value	844.7	145.7	52.1	25.1	14.3	9.0	6.1	4.3

FIGURE 8. Frequency Distribution of Node Visiting



Non-linear least square estimation was used to determine the values of parameters A and B in model $Y=A/X^B$. The result was A=844.691 and B=2.535. Expected values calculated by model $Y=844.691/X^{2.535}$ were listed at the bottom line of Table 15. The closeness of the expected values to the observed values indicates that the model fits the data very well. This closeness can also been seen from Figure 8 where the observed curve and the expected curve were plotted together. K-S test result further confirmed this conclusion. K-S value is 0.02 which is less than the critical value of 0.454

with n=8. A chi-square goodness of fit test was not conducted since it is not suitable for this kind of long tail distribution.

Although the Zipf distribution fits the general shape of the curve well, it was noticed that the fit is not so good at the tail end. Therefore, the most common discrete distributions, namely binomial distribution, Poisson distribution, and negative binomial distribution, were tried. Because the frequency of node visiting cannot be zero, the shifted version of these models were tested. The results are as follows.

The mean of the observed data is 1.36 and the variance 0.72. The shifted binomial distribution cannot be used, because the data has the property of $s^2 > \bar{X}-1$ which resulted in a $\hat{P}<0$. The parameter estimation for the shifted negative binomial distribution resulted in $\hat{P}=0.495$ and $\hat{V}=0.349$. The $\hat{V}$ was rounded up to the nearest integer 1, and the $\hat{P}$ was adjusted to 0.7372. The model passed the K-S test (K-S value is 0.044 which is less than the critical value of 0.454 with n=8). The shifted Poisson distribution also passed the K-S test (K-S value is 0.081, less than the critical value of 0.454). The expected value calculated from all these models which passed the goodness of fit test are listed in Table 17 for comparison.

Table 17 Comparison of Fitting of Different Models

X	1	2	3	4	5	6	7	8
observed Y	844	154	47	16	9	8	1	1
expected Y (Zipf)	844.7	145.7	52.1	25.1	14.3	9.0	6.1	4.3
expected Y (shifted Poisson)	756.2	269.6	48.0	5.7	0.5	0.04	0.002	0.0001
expected Y (shifted negative binomial)	796.2	209.2	55.0	14.5	3.8	0.998	0.262	0.069

Although all these three models passed the goodness of fit test, it is clear from Table 17 that the Zipf distribution fits best. The K-S values of these models also confirm this. The Zipf distribution has the lowest K-S value (0.02 compared to 0.044 for the shifted negative binomial distribution and 0.081 for the shifted Poisson distribution). Therefore, the Zipf distribution is the best model for the frequency distribution of node visiting.

Parameter A is the number of nodes that was visited once ($Y=A/1^B=A$). Parameter B indicates the steepness of the distribution curve. The larger the B value, the steeper the curve. In other words, the larger the B value, the larger the Y value discrepancy for the two adjacent X values. Parameter A and B are system dependent. The size of a database (total number of nodes in a database) can influence A value. Actually, parameter A also depends on the sample size. If twenty instead of fifty five searches were conducted in the experiment, the A value would be a lot smaller than 844. Parameter B could be influenced by the features of the system. For example, it is possible that the richness of the links (average number of links per node) will influence the B value. However, this is just a hypothesis. Further study can investigate the relationship between richness of links and frequency of node revisiting. The number of links could be another variable in the distribution model.

Nevertheless, the system dependent property of the parameters does not undermine the value of the model. As the first attempt of applying Zipf distribution to the hypertext path pattern, this study laid the background for further research. Furthermore, it can be conjectured that the frequency distribution of node visiting of any other hypertext system will follow the Zipf distribution, though with different parameter values. In other words, this model can be applied to other hypertext systems with re-estimation of parameters based on the system being used. This conjecture is based on the fact that Zipf's law, which was first discovered for English text, was found to be applicable to other languages and even indexing languages with differences in parameter values. Generally speaking, knowing the structure of the model is more important than knowing the concrete values of the parameters.

Chapter 6 Conclusions

The Markov model for hypertext search patterns proved to be a second-order one, that is, the past two moves influence the current move. Earlier moves have no influence on the current move.

The predominant search strategy was QUERYING (including search through Table of Contents, the Index, and keywords) -- DISPLAYING TITLE -- DISPLAYING ARTICLE -- PRINTING. It is clear that this search strategy is similar to that used in traditional Boolean logic type systems. NON-LINEAR BROWSING (including searching through citation or other highlighted terms in the text) was an auxiliary strategy. Thus, the keyword and Boolean search functions should be fully implemented in a hypertext system since they will be the major features used. Inadequate implementation of these functions will cause user dissatisfaction.

The users' search patterns and comments suggest that the system is, in general, easy to use (the help screen was rarely used), but the following features of the system should be improved:

(1) Faster keyword searching and full support of Boolean connectives should be implemented.

(2) The user's path history should be presented in a map form which shows the user's path and current position in relation to the global structure of the database. The current system provides only a linear list of nodes visited, and this is not particularly helpful when users are disoriented.

(3) The system should provide the option of displaying an article directly (in the current system, an article cannot be displayed directly without displaying the description of the article first). It would be more efficient to display the article immediately (one move) instead of displaying the title and the brief description first, and then the article (two moves). Mainly because a couple of lines of brief description of the article is not informative enough for users to make relevance judgement, they still have to see the article (84% chances indicated by the transition probability matrix).

Search patterns of different user groups were studied by comparing the corresponding transition probability matrices. Comparisons were made based on the following factors: gender, search experience, search task, and the subject's academic background. All the four comparisons revealed a significant difference between the groups being compared. The patterns of the difference are summarized below.

The difference between the two task groups is the strongest among the factors studied. For specific tasks, users tend to adopt the structured search pattern: keyword searching -- displaying title -- displaying article. For a general task, users tend to wander from node to node browsing without a structured path pattern. Because the strategy used by the specific task group can be realized very well in a Boolean logic-based system while the non-linear browsing uses more hypertext features, it can be said that the Boolean logic system is more suitable for specific search tasks and hypertext systems are more appropriate for general tasks. A system that has all the hypertext features and at the same time is well indexed and supports all the Boolean connectives will certainly be ideal for any task. The difference revealed in this comparison also has implication for user training.

Males are more likely than females to adopt the search strategy: keyword searching -- displaying title -- displaying article -- printing which is a typical search strategy in Boolean logic based systems. Females have a higher probability than males of continuous node revisiting and linear browsing. It is difficult to interpret this gender difference.

Highly experienced users tend to use hypertext features, moving from node to node non-linearly and later tracing back continuously while inexperienced users tend to apply more linear browsing. One possible interpretation of the difference is that experienced users feel more confident and relaxed in computerized retrieval systems, so that they can wander away from the current node freely and later trace back continuously without being too concerned about disorientation.

Although the statistical test conducted to study the factor of academic background indicated that there was a significant difference between users with technical background (science or engineering) and non-technical background (social science of humanities), the pattern of difference is not very clear and the small sample size for the technical background group (seven subjects) makes this test result unreliable.

The study also compared users' search options in terms of analytical searching vs. browsing. Effects of three factors (gender, search experience, and search task) were studied in a three-way analysis of variance. Search task influenced users' search behaviour. A general search task caused more browsing behaviour while a specific search task led to analytical search behaviour. Search experience alone did not affect search behaviour. There might be a gender difference among high experience users with males being more in favour of analytical searching than females. The conclusion about the gender difference here is subject to further research with a larger and more balanced sample.

Two frequency distribution models were developed and tested by the goodness-of-fit test to describe path patterns. The type of search task does not influence path lengths. The path length follows a shifted negative binomial distribution. The frequency of node visiting (how many nodes were visited once, how many twice, etc.) follows a Zipf distribution. Both models passed goodness-of-fit tests.

Although the Markov model has been used by earlier researchers to describe the search process in information retrieval systems, as reviewed the Chapter 3.1, this study is the first attempt to determine the order of such Markov models. With traditional information retrieval systems, numerous studies have been carried out to examine the factors that influence user performance. The factors that have been considered to affect user behaviour are mainly search experience and search task. Borgman (1986) also reported gender difference and academic background difference. However there was no detailed report as to how they differ in search patterns. With hypertext information retrieval systems, the factors that have been considered to influence search behaviour are for the most part the same as those for traditional systems. Nielsen (1989) summarized them as being individual differences among users and the effect of different tasks. Although there are fewer user studies in hypertext systems than in traditional systems, questions have been raised as to other possible factors that affect search behaviour, suitability of hypertext systems to different types of search tasks, and the indices for characterizing navigation (Barfield et al, 1990, Hendry, 1989, Marchionini et al, 1990). This study attempts to answer some of those questions by investigating the effect of gender, search task, search experience, and academic background. All these factors proved to influence search behaviour. The patterns of difference have been examined and summarized.

Every study has its limitations, and this study is no exception. It is necessary to discuss the study's limitations, because the conclusions stated above are subject to the conditions of the experiment and the data analysis techniques used. The experimental system used was the <u>Hyperties</u> version (IBM-PC) of <u>Hypertext on Hypertext</u>.

Compared to some other hypertext systems, which have a rich set of link types, Hyperties supports only a single type of link, one that points from a text string within an article to another article. When more links are available, users will have more free choice and therefore may have a different search pattern. For example, the frequency distribution of node visiting could be different from that described in Chapter 5.5. The database Hypertext on Hypertext is fairly small compared to most real databases. It is possible that the size of the database influenced path length, and a larger database would lead to longer path lengths. However, it is conjectured that the difference in the system and database will influence only the parameters of the models developed in this study but not their structure. That is, the frequency distribution of node visiting will still have a hyperbolic shape and the path length distribution will still be in a bell shape. This is only a hypothesis, and needs to be tested in future studies.

Another major limitation of this study is the subjects used in the experiment. Since most subjects are library and information science students whose search behaviour could be different from others, conclusions in this study cannot be generalized to other populations without caution. Besides, the discrepancy in sample sizes (e.g. between technical and non-technical groups), the difference in motivation of conducting the search, and the difference in the training environment between the volunteers and the students can all bias the models developed and the conclusions reached in this study.

In spite of these limitations, the probabilistic models presented and tested in this study can help us better understand users' search behaviour and the search process involved in hypertext information retrieval systems. They provide valuable information for evaluating a system's existing operation and for refining future design. They also provide a background for the examination of systems via mathematical models and simulation studies. Finally, they provide a description of how individual differences can affect hypertext search patterns.

APPENDIX 1. Description of Search Tasks

Type 1 -- General Task

Imagine that you have been charged with writing a one-page encyclopedia article on "HYPERTEXT" for the Encyclopedia of Library and Information Science. However, the only source material you can use must come from the ACM Hypertext on Hypertext database. Make a shopping list of information you will want (origin of term, inventor, history, definition, advantages and disadvantages, areas of application, problem areas, future trend, bibliographic references, etc.) and search the database to try and get as much of it as possible. Then write the concise article (using only material found in Hypertext on Hypertext).

Type 2 -- Specific Task

Each subject conducted a search on one of the following topics:

1. Briefly discuss the history of hypertext systems.
2. Is hypertext representation always appropriate? Why?
3. Is converting text to hypertext always appropriate? Why?
4. What are the problems associated with hypertext systems?

5. How does a hypertext "document" differ from a conventional paper document?
6. What events during the past years caused the interest in hypertext to accelerate sharply?
7. How should the system support browsing?
8. Classify hypertext systems.
9. What hardware is required to run Hyperties software.
10. What size should a node be?
11. What types of links should there be?
12. What measures can be used to reduce disorientation ?
13. Describe the future of hypertext systems.

APPENDIX 2. Questionnaire

1. Subject No. ______________________

2. Gender: male female

3. Academic background (please circle one of the following)

   Social science

   Humanities

   Engineering

   Science

4. Experience with online (not OPACs) or CD-ROM system. Circle from 1 (no experience) to 5 (expert).

   1 2 3 4 5

5. Search topic:

APPENDIX 3 SECOND-ORDER TRANSITION PROBABILITY MATRICES

Condition: move 1 is QUERYING

move 3 / move 2	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
querying	0.266	0.709	0.000	0.025	SZ	SZ	SZ	0.000
displaying title	0.015	0.084	0.010	0.011	0.874	SZ	SZ	0.006
printing	0.000	0.667	SZ	0.333	0.000	0.000	0.000	0.000
node revisiting	0.440	0.240	0.000	0.080	SZ	0.080	0.040	0.120
displaying article	SZ	SZ	SZ	SZ	SZ	SZ	SZ	SZ
non-linear browsing	SZ	SZ	SZ	SZ	SZ	SZ	SZ	SZ
linear browsing	SZ	SZ	SZ	SZ	SZ	SZ	SZ	SZ
getting help	0.643	0.071	0.000	0.286	SZ	0.000	0.000	0.000

APPENDIX 3 SECOND-ORDER TRANSITION PROBABILITY MATRICES (continued)

Condition: move 1 is DISPLAYING TITLE

move 3 / move 2	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
querying	0.250	0.625	0.000	0.063	SZ	SZ	SZ	0.063
displaying title	0.024	0.492	0.000	0.000	0.484	SZ	SZ	0.000
printing	0.143	0.429	SZ	0.000	0.429	0.000	0.000	0.000
node revisiting	0.571	0.143	0.000	0.000	SZ	0.143	0.143	0.000
displaying article	0.215	0.195	0.249	0.094	SZ	0.126	0.103	0.018
non-linear browsing	SZ	SZ	SZ	SZ	SZ	SZ	SZ	SZ
linear browsing	SZ	SZ	SZ	SZ	SZ	SZ	SZ	SZ
getting help	0.250	0.000	0.000	0.750	SZ	0.000	0.000	0.000

APPENDIX 3 SECOND-ORDER TRANSITION PROBABILITY MATRICES (continued)

Condition: move 1 is PRINTING

move 2 \ move 3	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
querying	0.080	0.860	0.030	0.010	SZ	SZ	SZ	0.020
displaying title	0.010	0.019	0.019	0.000	0.952	SZ	SZ	0.00
printing	SZ	SZ	SZ	SZ	SZ	SZ	SZ	SZ
node revisiting	0.149	0.040	0.089	0.168	SZ	0.257	0.257	0.040
displaying article	0.000	0.000	1.000	0.000	SZ	0.000	0.000	0.000
non-linear browsing	0.103	0.103	0.231	0.154	SZ	0.333	0.077	0.000
linear browsing	0.096	0.036	0.434	0.120	SZ	0.096	0.205	0.012
getting help	0.375	0.000	0.125	0.500	SZ	0.000	0.000	0.000

APPENDIX 3 SECOND-ORDER TRANSITION PROBABILITY MATRICES (continued)

Condition: move 1 is NODE REVISITING

move 3 / move 2	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
querying	0.075	0.805	0.015	0.075	SZ	SZ	SZ	0.030
displaying title	0.038	0.038	0.000	0.000	0.925	SZ	SZ	0.000
printing	0.148	0.185	SZ	0.148	0.000	0.185	0.296	0.037
node revisiting	0.261	0.120	0.022	0.250	SZ	0.120	0.185	0.043
displaying title	SZ	SZ	SZ	SZ	SZ	SZ	SZ	SZ
non-linear browsing	0.117	0.000	0.194	0.437	SZ	0.204	0.049	0.000
linear browsing	0.055	0.022	0.319	0.308	SZ	0.099	0.176	0.022
getting help	0.313	0.000	0.031	0.656	SZ	0.000	0.000	0.000

APPENDIX 3 SECOND-ORDER TRANSITION PROBABILITY MATRICES (continued)

Condition: move 1 is DISPLAYING ARTICLE

move 3 / move 2	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
querying	0.112	0.849	0.006	0.022	SZ	SZ	SZ	0.011
displaying title	0.000	0.069	0.000	0.006	0.919	SZ	SZ	0.006
printing	0.343	0.338	SZ	0.052	0.000	0.100	0.157	0.010
node revisiting	0.474	0.269	0.026	0.141	SZ	0.038	0.038	0.013
displaying article	SZ	SZ	SZ	SZ	SZ	SZ	SZ	SZ
non-linear browsing	0.151	0.047	0.217	0.302	SZ	0.189	0.075	0.019
linear browsing	0.059	0.094	0.353	0.235	SZ	0.094	0.153	0.012
getting help	0.214	0.071	0.000	0.643	SZ	0.071	0.000	0.000

APPENDIX 3 SECOND-ORDER TRANSITION PROBABILITY MATRICES (continued)

Condition: move 1 is NON-LINEAR BROWSING

move 2 \ move 3	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
querying	0.043	0.870	0.000	0.087	SZ	SZ	SZ	0.000
displaying title	0.091	0.091	0.000	0.000	0.818	SZ	SZ	0.000
printing	0.100	0.071	SZ	0.529	0.000	0.143	0.129	0.029
node revisiting	0.217	0.035	0.052	0.139	SZ	0.478	0.043	0.035
displaying article	SZ	SZ	SZ	SZ	SZ	SZ	SZ	SZ
non-linear browsing	0.074	0.013	0.074	0.134	SZ	0.617	0.054	0.034
linear browsing	0.103	0.034	0.517	0.069	SZ	0.103	0.172	0.000
getting help	0.125	0.000	0.000	0.625	SZ	0.125	0.000	0.125

APPENDIX 3 SECOND-ORDER TRANSITION PROBABILITY MATRICES (continued)

Condition: move 1 is LINEAR BROWSING

move 2 \ move 3	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
querying	0.138	0.759	0.000	0.034	SZ	SZ	SZ	0.069
displaying title	0.000	0.222	0.000	0.000	0.778	SZ	SZ	0.000
printing	0.128	0.136	SZ	0.384	0.000	0.040	0.288	0.024
node revisiting	0.132	0.013	0.053	0.263	SZ	0.000	0.513	0.026
displaying article	SZ	SZ	SZ	SZ	SZ	SZ	SZ	SZ
non-linear browsing	0.094	0.000	0.219	0.438	SZ	0.063	0.156	0.031
linear browsing	0.116	0.058	0.261	0.232	SZ	0.058	0.275	0.000
getting help	0.500	0.000	0.000	0.500	SZ	0.000	0.000	0.000

APPENDIX 3 SECOND-ORDER TRANSITION PROBABILITY MATRICES (continued)

Condition: move 1 is GETTING HELP

move 2 \ move 3	querying	displaying title	printing	node revisiting	displaying article	non-linear browsing	linear browsing	getting help
querying	0.214	0.643	0.000	0.036	SZ	SZ	SZ	0.107
displaying title	0.333	0.000	0.000	0.333	0.333	SZ	SZ	0.000
printing	1.000	0.000	SZ	0.000	0.000	0.000	0.000	0.000
node revisiting	0.174	0.109	0.109	0.087	SZ	0.109	0.087	0.326
displaying article	SZ	SZ	SZ	SZ	SZ	SZ	SZ	SZ
non-linear browsing	0.000	0.000	0.000	0.000	SZ	1.000	0.000	0.000
linear browsing	0.000	0.000	0.000	0.000	SZ	0.000	0.000	0.000
getting help	0.000	0.000	0.000	0.000	SZ	0.000	0.000	0.000

REFERENCES

Barfield, W., Haselkorn, M., & Weatbrook, C. (1990). Information Retrieval with a Printed User's Manual and with Online HyperCard Help. Technical Communication, 37(1), 23-27.

Beheshti, J., & Tague, J. (1984). Morse's Markov Model of Book Use Revisited. Journal of the American Society for Information Science, 35(5), 259-267.

Bernstein, M. (1988). The Bookmark and the Compass: Orientation Tools for Hypertext Users. SIGOIS Bulletin (USA), 9(4), 34-45.

Billingsley, P. (1961). Statistical Methods in Markov Chains. Annals of Mathematical Statistics, 32(1), 1961, 12-41.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). Discrete Multivariate Analysis: Theory and Practice. Cambridge, U.S.A.: MIT Press.

Bookstein, A. (1977). [Review of Applications of Operations Research Models to Libraries; A Case Study of the Use of Monographs in the Francis A. Countway Library of

Medicine, Harvard University]. Journal of Library Automation, 10(2), 197-198.

Borgman, C. L. (1982). Online Monitoring of Users of the OHIO State University Online Catalog: Methodological Issues and Results. Proceedings of the 45th ASIS Annual Meeting, 35-36.

Borgman, C. L. (1986). The User's Mental Model of an Information Retrieval System: An Experiment on a Prototype Online Catalog. International Journal of Man-Machine Studies, 24(1), 47-64.

Borgman, C. L. (1989). All Users of Information Retrieval Systems Are Not Created Equal: An Exploration into Individual Differences. Information Processing and Management, 25(3), 237-251.

Brueggeman, P. (1989). Observation and Playback of CD-ROM Searches with Total Recall. CD-ROM Librarian, 4(6), 12-16.

Campagnoni, F. R. & Ehrlich, K. (1989). Information Retrieval Using A Hypertext-Based Help System. Proceedings of the 12 Annual International ACM SIGIR Conference on Research and Development in Information

Retrieval, 212-220.

Chapman, J. L. (1981). A State Transition Analysis of Online Information-Seeking Behaviour. Journal of the American Society for Information Science, 32(5), 325-333.

Chen, C. (1976). Applications of Operations Research Models to Libraries; A Case Study of the Use of Monographs in the Francis A. Countway Library of Medicine, Harvard University. Cambridge, U.S.A.: M.I.T. Press.

Coady, R. P. (1981). The Applicability of Markov Models to the Circulation of Social-Science Monographs in a Large Academic Library. Ph.D. Dissertation, Indiana University.

Cooper, M. D. (1973). A Simulation Model of an Information Retrieval Systems. Information Storage and Retrieval, 9, 13-32.

Dumais, S. T. (1988). Textual Information Retrieval. In Handbook of Human-Computer Interaction, edited by M. Helander. North-Holland: Elsevier Science Publishers B. V.

Edward, D. M. & Hardman, L. (1989). 'Lost in Hyperspace':

Cognitive Mapping and Navigation in a Hypertext Environment. In Hypertext: theory into practice, edited by R. McAleese. Osney Mead, Oxford: Blackwell Scientific Publications Ltd.

Egan, D. et al. (1989). Formative Design-Evaluation of SuperBook. ACM Transactions on Information Systems, 7(1), 30-57.

Fenichel, C. H. (1981). Online Searching: Measures that Discriminate among Users with Different Types of Experiences. Journal of the American Society for Information Science, 32(1), 23-32.

Fienberg, S. E. (1980). The Analysis of Cross-Classified Categorical Data. Cambridge, U.S.A.: MIT Press, 2nd edition.

Foss, C. L. (1989). Tools for Reading and Browsing Hypertext. Information Processing and Management, 25(4), 407-418.

Foster, J. (1986). Online Searching: A Five Star Review of Research. The Canadian Journal of Information Science, 11(1), 1-17.

Franklin, C. (1989). Hypertext Defined and Applied. Online, 13(3), 37-49.

Gary, S. H., & Shasha, D. (1989). To Link or not to link? Empirical Guidance for the Design of Nonlinear Text Systems. Behaviour Research Methods, Instruments, and Computers, 21(2), 326-333.

Glushko, R. J. (1990). Using Off-the-Shelf Software to Create a Hypertext Electronic Encyclopedia. Technical Communication, 37(1), 28-33.

Goffman, W. (1971). Mathematical Method for analyzing the Growth of a Scientific Discipline. Journal of the Association for Computing Machinery, 18(2), 173-185.

Gomez, L. M., Egan, D. E., Wheeler, E. A., Sharma, D. K., & Gruchacz, A. M. (1983). How Interface Design Determines Who Has Difficulty Learning to Use a Text Editor. Proceedings of CHI Conference on Human-computer Interaction, 176-181.

Griffiths, J. M. (1975). Index Term Input to IR Systems. Journal of Documentation, 31, 185-190.

Griffiths, J. M. (1977). The Computer Simulation of

Information Retrieval Systems. Ph.D. Dissertation, London, University College.

Halasz, F. (1987). Reflections on NoteCards: Seven issues for the next generation of hypermedia systems. Proceedings of HyperTEXT'87, 345-366.

Harris, M. A. (1986). Sequence Analysis of Moves in Online Searching. The Canadian Journal of Information Science, 11(2), 35-56.

Hendry, D. G. (1989). The Relationship Between Navigation and Comprehension in a Hypertext Environment. Masters thesis, Department of Computing and Information Science, University of Guelph, Canada.

Hillier, F. S., & Lieberman, G. J. (1972). Introduction to Operations Research. San Francisco: Holden-Day.

Johnson, N. L., & Kotz, S. (1969). Discrete Distributions. Boston, U.S.A.: Houghton Miflin Company.

Judd, C. M. & McClelland, G. H. (1989). Data Analysis: A Model-Comparison Approach. San Diego: Harcourt Brace Jovanovich.

Keppel, G. & Saufley, W. H. Jr. (1980). Introduction to Design and Analysis - A Student's Handbook. San Francisco: W. H. Freeman and Company.

Knoke, D., & Burde, P. J. (1980). Log-Linear Models. London, England: Sage Publications.

Mandelbrot, B. (1953). An Information Theory of the Statistical Structure of Language. In W. Jackson (Ed.), Communication Theory, (pp. 486-502). London: Butterworths.

Mantel, M. M. (1982). Disorientation Behaviour in Person-computer Interaction. Ph.D. thesis, University of Southern California.

Marchionini, G. (1987). An Invitation to Browse: Designing Full-text Systems for Novice Users. Canadian Journal of Information Science, 12(3-4), 69-79.

Marchionini, G. (1989). Information-Seeking Strategies of Novices Using a Full-Text Electronic Encyclopedia. Journal of the American Society for Information Science, 40(1), 54-66.

Marchionini, G., Lin, X., & Dwiggins, S. (1990). Effects of

Search and Subject Expertise on Information Seeking in a Hypertext Environment. Proceedings of the 53rd Annual Meeting of the American Society for Information Science, Toronto, Canada, Nov. 4-8, 1990, 129-142.

Marchionini, G., & Shneiderman, B. (1988). Finding Facts vs. Browsing Knowledge in Hypertext Systems. Computer, 21(1), 70-80.

McAleese, R. (1989). Navigation and Browsing in Hypertext. In Hypertext: theory into practice, edited by McAleese, R., Osney Mead, Oxford: Blackwell Scientific Publications Ltd.

Morse, P. M. (1968). Library Effectiveness: A Systems Approach. Cambridge, U.S.A.: M.I.T. Press.

Nelson, M. J. (1981). Probabilistic Models for the Simulation of Bibliographic Retrieval Systems. Ph.D. Dissertation, The University of Western Ontario.

Nelson, M. J., & Tague, J. M. (1985). Split Size-Rank Models for the Distribution of Index Terms. Journal of the American Society for Information Science, 36(5), 283-296.

Nielsen, J. (1989). The Matters that Really Matter for Hypertext Usability. Hypertext'89: Proceedings of the Second ACM Conference on Hypertext, 239-248.

Nielsen, J. (1990). The Art of Navigating through Hypertext. Communication of ACM, 33(3), 296-310.

Pao, M. L., & McCreery, L. (1986). Bibliometric Application of Markov Chains. Information Processing & Management, 22(1), 7-17.

Penniman, W. D. (1975a). Rhythms of Dialogue in Human-Computer Conversation. Ph.D. Dissertation, Ohio State University.

Penniman, W. D. (1975b). A Stochastic Process Analysis of On-Line User Behaviour. Information Revolution: Proceedings of the 38th ASIS Annual Meeting, 147-148.

Rice, R. E., & Borgman C. L. (1983). The Use of Computer-Monitored Data in Information Science and Communication Research. Journal of the American Society for Information Science, 34(4), 247-256.

Schultz, C. K., Schwartz, P. D., & Steinberg, L. (1961). A Comparison of Dictionary Use Within Two Information

Retrieval Systems. American Documentation, 12, 247-253.

Tague, J., & Ajiferuke, I. (1987). The Markov and the Mixed-Poisson Models of Library Circulation Compared. Journal of Documentation, 43(3), 212-235.

Tolle, J. E. (1983a). Current Utilization of Online Catalogs: Transaction Log Analysis -- Final Report to the Council on Library Resources, Volume I of three volumes.

Tolle, J. E. (1983b). Transaction Log Analysis, Online Catalogs. Proceedings of the 6th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 147-160.

Tolle, J. E. (1984a). Monitoring and Evaluation of Information Systems Via Transaction Log Analysis. Proceedings of the 3rd Joint BCS and ACM Symposium, 247-258.

Tolle, J. E. (1984b). Monitoring Usage of Online Information Systems: NLM CATLINE Database. Proceedings of the 47th ASIS Annual Meeting, 126-129.

Tolle, J. E., & Hah, S. (1985). Online Search Patterns: NLM

CATLINE Database. Journal of the American Society for Information Science, 36(2), 82-93.

Vigil, P. J. (1988). Online Retrieval: Analysis and Strategy. New York: John Wiley & Sons.

Walker, J. H., & Young, E., & Mannes, S. A Case Study of Using a Manual Online. personal communication.

Zellwger, P. T. (1989). Scripted Documents: A Hypermedia Path Mechanism. Hypertext'89: Proceedings of the Second ACM Conference on Hypertext, 1-14.

Zipf, G. K. (1949). Human Behaviour and the Principle of Least Effort. Cambridge, Mass.: Addison-Wesley.

Zunde, P., & Slamecka, V. (1971). Predictive Models of Scientific Progress. Information Storage and Retrieval, 7(3), 103-109.